GPT-6 as a Universal Foundation Model for Protein Biology: Beyond Specialized Models

Masatoshi Uehara[1], Edward Zhou[1], Joshua Park[1], Morgan Griffiths[1],

Austin Clyde[1], Suyash Shringarpure[1], Yuan Yuan[1], Kevin Rao[1],

Anton Bushuiev[2], Michael Yu[2], Kathryn M. Collins[3], Jesse Hendricks[3], Puwanat Sangkhapreecha[3], Lee-Shin Chu[3], Sharrol Bachas[3], Joy Jiao[11]

[1] Life Science Team, OpenAI

[2] Contractor, OpenAI

[3] Onava AI

November 24, 2026

Abstract

Specialized models have transformed protein structure prediction and design, yet incorporating new molecular constraints often requires modifying the software or retraining the model. Here we show that general-purpose language models in the GPT-6 family can perform diverse protein modeling and design tasks using only reasoning and code, with natural language as a common means of specifying molecular constraints. We experimentally evaluated a pre-release version of GPT-6.2 on de novo backbone generation and inverse folding. GPT-designed backbones paired with SolubleMPNN sequences yielded minibinders against all five protein targets tested. Inverse folding of nonpublic target–binder backbones yielded binders whose sequences differed extensively from the withheld parental sequences. Antibody inverse folding and therapeutic antibody redesign likewise yielded variants that retained binding despite substantial substitutions in their complementarity-determining regions. Across computational benchmarks, GPT-6 models often matched or exceeded specialized baselines in structure and sequence design, ligand docking and variant effect prediction, while remaining substantially less accurate in global fold prediction. By changing the prompt alone, we could specify binding interfaces, prescribe protein's secondary structures, control sequence divergence or constrain ligand poses. This steerability allows researchers to incorporate new structural information and revise design objectives within the same model, without further training. Together, these results demonstrate broad intrinsic protein capabilities in general-purpose language models and support a path toward programmable molecular design guided by natural-language instructions and biological reasoning.

[1]Corresponding authors: masatoshiuehara@openai.com, yjiao@openai.com The OpenAI Life Science Team leads this project. Onava contributed by conducting a protein design campaign in the wet lab, including experimental design

Contents

1 Introduction **3**

2 Capability Analysis: Protein Structure Design **6**

3 Capability Analysis: Protein Sequence Design **17**

4 Capability Analysis: Structure Prediction **30**

5 Capability Analysis: Variant Effect Prediction **39**

6 How Does GPT Address Problems in Protein Biology? **40**

7 Conclusion — Towards Programmable Molecular Design **47**

8 Methods **48**

Acknowledgements **50**

References **50**

Appendix **60**

A Details of Figure 2 **60**

B Experimental Methods **61**

C Generation and Experimental Validation of Parental VHH Binders **63**

D Additional Analyses of Experimental Studies **64**

E Further Examples of How Models Solve Protein Biology **79**

1 Introduction

Recent advances in specialized artificial intelligence (AI) models have transformed protein structure prediction and computational design, opening new opportunities for biological discovery and therapeutic development. These advances span several complementary tasks, including the generation of protein structures, sequence design, and the prediction of structural and functional properties. For example, RFdiffusion (Watson et al. 2023) enables the generation of protein backbones subject to structural constraints, while ProteinMPNN (Dauparas et al. 2022) designs amino acid sequences compatible with specified backbone geometries. Complementing these generative approaches, AlphaFold (Jumper et al. 2021) predicts protein structures from sequence, and protein language models such as the ESM family of models (Rives et al. 2021) learn sequence representations that can be leveraged for downstream predictive tasks. Together, these models provide complementary capabilities that can be integrated into computational protein design workflows.

Increasingly capable LLMs can serve as agents that orchestrate these specialized models and tools. Recent examples include the Virtual Lab, which developed a nanobody-design pipeline combining ESM, AlphaFold-Multimer, and Rosetta (Swanson et al. 2025); Biomni, which integrates biomedical tools and databases through code-based execution (Huang et al. 2026); and Chaperone, which enables LLM agents to coordinate RFdiffusion3, ProteinMPNN, AlphaFold 3, and other protein-design tools through standardized interfaces (Butcher, Chen, et al. 2026). An Anthropic technical report described a binder-design workflow in which Claude used an approximately 16,000-word prompt to coordinate a suite of protein-design and structure-prediction tools, with subsequent experimental validation of binding (Claude Science and Shanehsazzadeh 2026).

Despite expert agentic orchestration, designs resulting from tool-based workflows remain constrained by the capabilities and steerability of the specialized tools.

- Limited Capabilities: In specialized protein models, generalization beyond familiar sequence, structural, and functional contexts remains a recurring challenge. Protein–ligand cofolding accuracy depends strongly on similarity to training examples (Škrinjar et al. 2026), and models can retain familiar poses despite perturbations to binding-site chemistry that should alter molecular recognition (Masters et al. 2025). Controlled tests of an AlphaFold2-derived model also reveal difficulty recovering alternative conformations absent from its training set (Schafer and Porter 2025). Furthermore, generalization beyond the tasks for which these models were originally designed remains limited. While representations learned by models such as ESM and AlphaFold are widely recognized as useful for predicting downstream properties (e.g., binding affinity and developability) through zero-shot or few-shot learning, their success in these tasks remains limited. Indeed, sequence-to-function benchmarks expose sensitivity to unseen substitutions, mutation positions, and protein scaffolds (Didi et al. 2026), while predicting improved function beyond the fitness range observed during training remains difficult across multiple datasets and models (Gelman et al. 2025).
- Steerability: The protein design community has made significant efforts to make protein generation steerable through bespoke conditioning mechanisms for many molecular contexts or structural constraints, such as those used in RFdiffusion3 (Butcher, Krishna, et al. 2026) and Proteina (Geffner et al. 2025), or through test-time conditioning, as demonstrated by Chroma (Ingraham et al. 2023). However, these conditioning approaches still have substantial limitations and can compromise generative fidelity. Natural language offers the most intuitive and informative way to incorporate relevant conditions: molecular contexts, including binding inter-

faces, binder shapes, and any conditioning information expressed in terms of molecular structures and shapes (DNA, RNA, and small molecules), post-translational modifications (PTMs) such as accessibility-altering glycans near a proposed interface (Adams et al. 2025; Greisen et al. 2023); experimental conditions; prior scientific findings; and expert knowledge within and across scientific fields.

Motivated by the bitter lesson (Sutton 2019) and the potential for new capabilities to emerge through scaling and reasoning, we pursued the central question of whether a single general-purpose LLM can acquire protein modeling capabilities without relying on specialized models, and what advantages such unification might offer. We use "intrinsic protein capabilities" to describe an LLM's ability to analyze and design proteins using only basic coding tools, without access to pretrained specialized models (Figure 1). This was our standard evaluation setting for GPT in the benchmarks in Sections 2–4. Where indicated, we used specialized models to score or filter GPT's outputs after submission. This paradigm could thus address the two challenges outlined above: the limited capabilities of specialized models and their limited steerability.

Here we show that models in the GPT-6 family have intrinsic protein design capabilities that allow them to overcome several limitations of specialized models. They combine extensive biological knowledge with strong reasoning skills in mathematics, coding (OpenAI 2026), and physics (Craig 2026). They achieve state-of-the-art performance across a wide range of protein design tasks and generalize to problems that require more than recall.

Natural-language instructions can incorporate biological context and experimental constraints. For example, an LLM could query UniProt (UniProt Consortium 2025) for post-translational modification (PTM) annotations and supporting evidence, assess their relevance to the target's biological context and proposed binding interface, and use this information to refine hotspot selection or candidate designs.

Main Results

Across the studies below, GPT-6.x (GPT-6, GPT-6.2 and more advanced internal models) often approached or exceeded specialized models' performance, whereas GPT-5.6 often lagged far behind. Figure 2 summarizes performance across tasks. Sections 2–5 describe each task and its results. Section 6 illustrates how the models combine biological knowledge with custom code to propose, evaluate, and iteratively refine candidate structures and sequences. Section 8 provides the evaluation settings, scoring criteria, and procedures for selecting designs for experimental testing.

Protein structure design: These tasks include de novo minibinder design against diverse molecular targets, including proteins, small molecules, DNA, and RNA, where models generate binder structures conditioned on target structures. Such capabilities have important applications in drug design. Other tasks include motif scaffolding, which supports the design of functional proteins such as enzymes and binders by constructing protein scaffolds around predefined structural motifs, and fold-conditioned protein generation based on CATH classifications. Across these tasks, we demonstrate that GPT-6.x can achieve performance comparable to specialized models such as RFdiffusion, RFdiffusion3, and Proteina.

Protein sequence design (i.e., inverse folding): These tasks involve designing protein and antibody sequences compatible with given backbone structures. Inverse folding is widely used to generate sequence-diverse protein variants, providing a foundation for functional optimization such as pro-

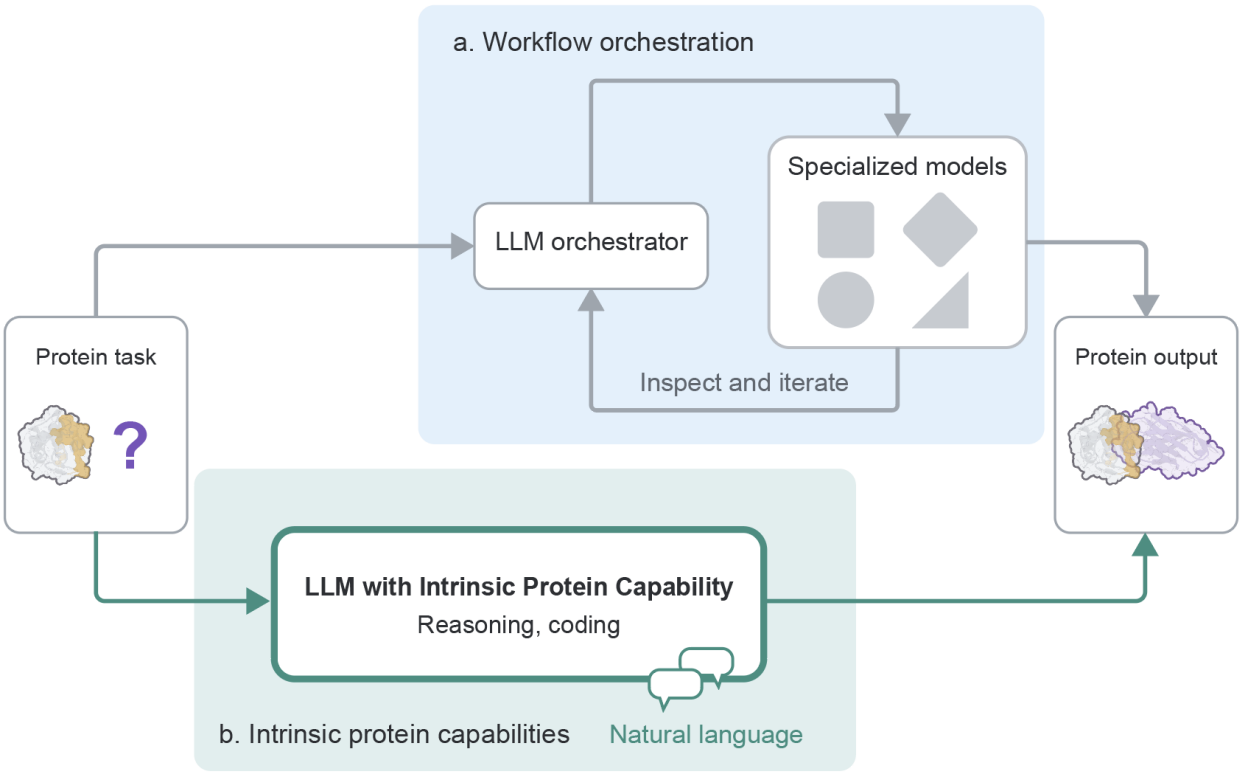


Figure 1.
Conceptual comparison of two approaches to protein modeling and design. (a) In a workflow-orchestration approach, an LLM interprets a protein task, calls specialized protein models, inspects their outputs, and iterates on the workflow. The specialized-model box represents tools for generating structures, assigning sequences, and predicting protein conformations. (b) An LLM with intrinsic protein capabilities uses its own reasoning and coding to produce protein outputs, with natural-language instructions steering the task. Here, intrinsic capability refers to candidate generation and analysis without calling pretrained specialized protein models.

tease (Krasnow et al. 2026), membrane proteins (Goverde et al. 2024), affinity maturation in antibody design (Shanker et al. 2024). We demonstrate that GPT-6.x can match the performance of specialized models such as ProteinMPNN (Dauparas et al. 2022) and AbMPNN (Dreyer et al. 2023).

Structure prediction: Structure prediction is a fundamental task in structural biology and drug design. It includes predicting the three-dimensional structure of a protein monomer from its amino acid sequence. State-of-the-art models such as AlphaFold 3 (Abramson et al. 2024) can also predict the structures of complexes involving multiple protein chains, DNA, RNA, and small molecules. In this work, to characterize performance in more detail, we evaluated four tasks separately (monomer backbone prediction, protein–protein docking, protein–ligand docking, and side-chain packing). While GPT-6.x still substantially trails state-of-the-art specialized models in global monomer structure prediction, it achieves strong performance in docking and side-chain packing, comparable to specialized models such as DiffDock (Corso et al. 2023, 2024) and FlowPacker (Lee and Kim 2025). Furthermore, in local structure prediction tasks, such as antibody CDR loop prediction, GPT-6.x approaches the performance of state-of-the-art specialized models.

Variant effect prediction: These tasks involve ranking protein variants relative to a wild-type sequence according to their functional properties, a central problem in protein engineering such as binding, pathogenicity, and regulatory activity. We benchmark GPT-6.x across diverse experimental assays measuring properties such as thermostability, binding and expression. GPT-6.x demonstrates strong intrinsic capabilities, achieving competitive performance against state-of-the-art zero-shot baselines using ESM2 (Lin et al. 2023).

Wet-Lab Validation

We tested a pre-release version of GPT-6.2 at two stages of minibinder design. In one campaign, GPT generated backbones and SolubleMPNN (Goverde et al. 2024) designed the sequences (Section 2.5). In another, GPT designed sequences from supplied backbones, yielding 123 screening hits across five targets (Section 3.3.1). Antibody screens also yielded hits when GPT designed loops without the parental loop sequences and when it redesigned supplied sequences (Sections 3.3.2–3.3.3). All campaigns used specialized structure prediction models to select candidates for testing (Section 8.3).

2 Capability Analysis: Protein Structure Design

Designing a protein to bind a specified site requires determining both its fold and its pose relative to the target. Specialized models such as RFdiffusion (Watson et al. 2023) and BoltzGen (Stark et al. 2025) address this problem by generating protein structures that satisfy specified geometric constraints. We asked whether GPT could generate the backbone coordinates itself, given structural data and a description of the design task in natural language. In our computational evaluations, models in the GPT-6 family matched the performance of specialized models on minibinder design for both protein and non-protein targets, motif scaffolding, and CATH-conditioned design.

2.1 De Novo Minibinder Design Against Protein Targets

Dataset and Settings. The task is to generate binder–target complex backbones given the structures of the protein targets. We used a set of commonly studied targets from prior binder design studies. By default, we included binding hotspot residues in the prompt to specify the desired binding interface. We evaluated GPT in the standard setting. The model further did not have access to the internet. We used RFdiffusion as a specialized baseline. We set binder lengths to 80–120 residues.

Evaluation. We counted a backbone as successful if at least one refold reproduced both the binder fold and its placement on the target and met the interface confidence threshold (Section 8.2).

2.1.1 Results

Our internal models generally achieve higher in silico success rates than RFdiffusion across the evaluated targets (Figure 3). These results suggest that GPT can intrinsically generate structurally plausible protein binders. We showcase several computationally successful binder designs in Figure 3 and further validate these capabilities through wet-lab experiments in Section 2.5.

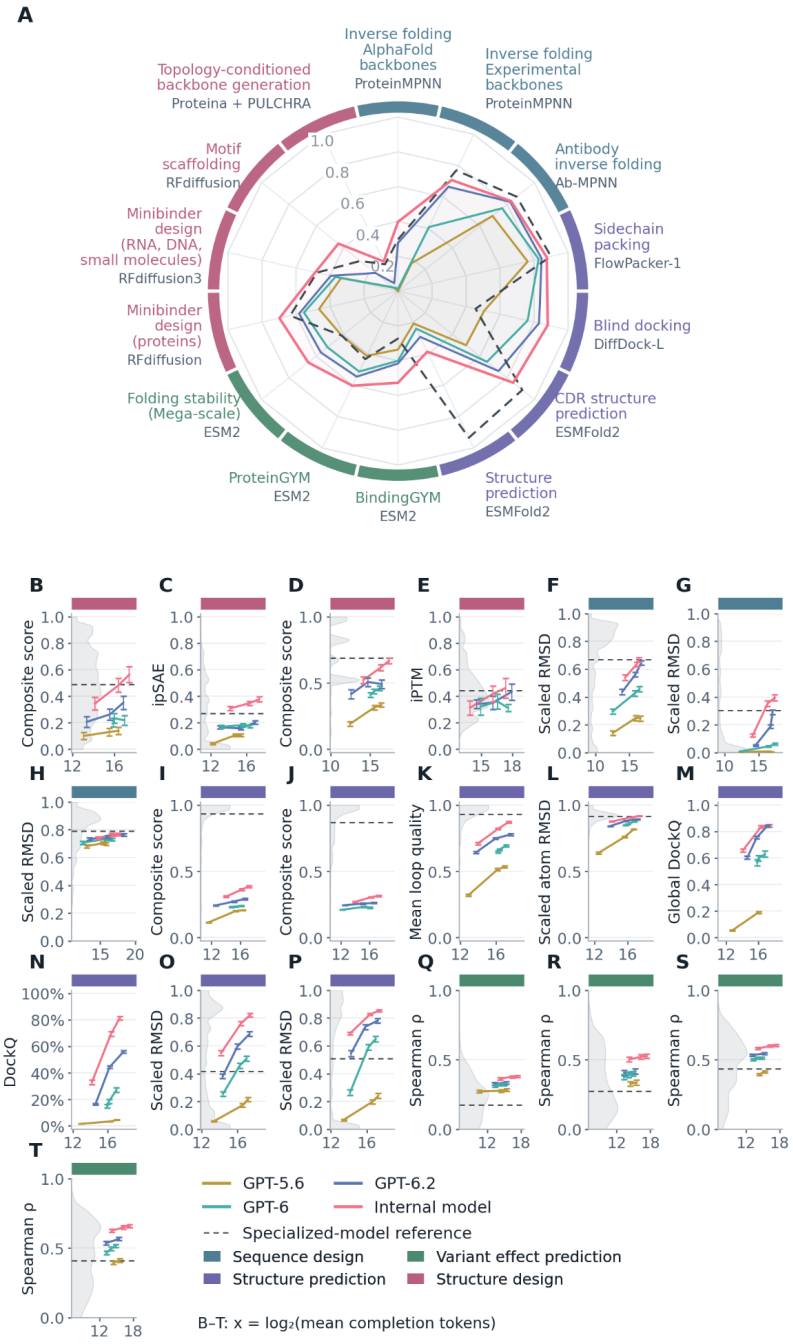


Figure 2. Performance across biological tasks. A compares performance at the highest evaluated budgets; B–T show how performance changes with mean generated completion tokens. The horizontal axis is $\log_2$(mean tokens), so one unit denotes twice as many tokens. Colors identify models and the bars identify task categories. Dashed horizontal lines mark specialized-model references. Gray distributions show specialist performance. Higher values indicate better performance within each task. Details in Appendix A.

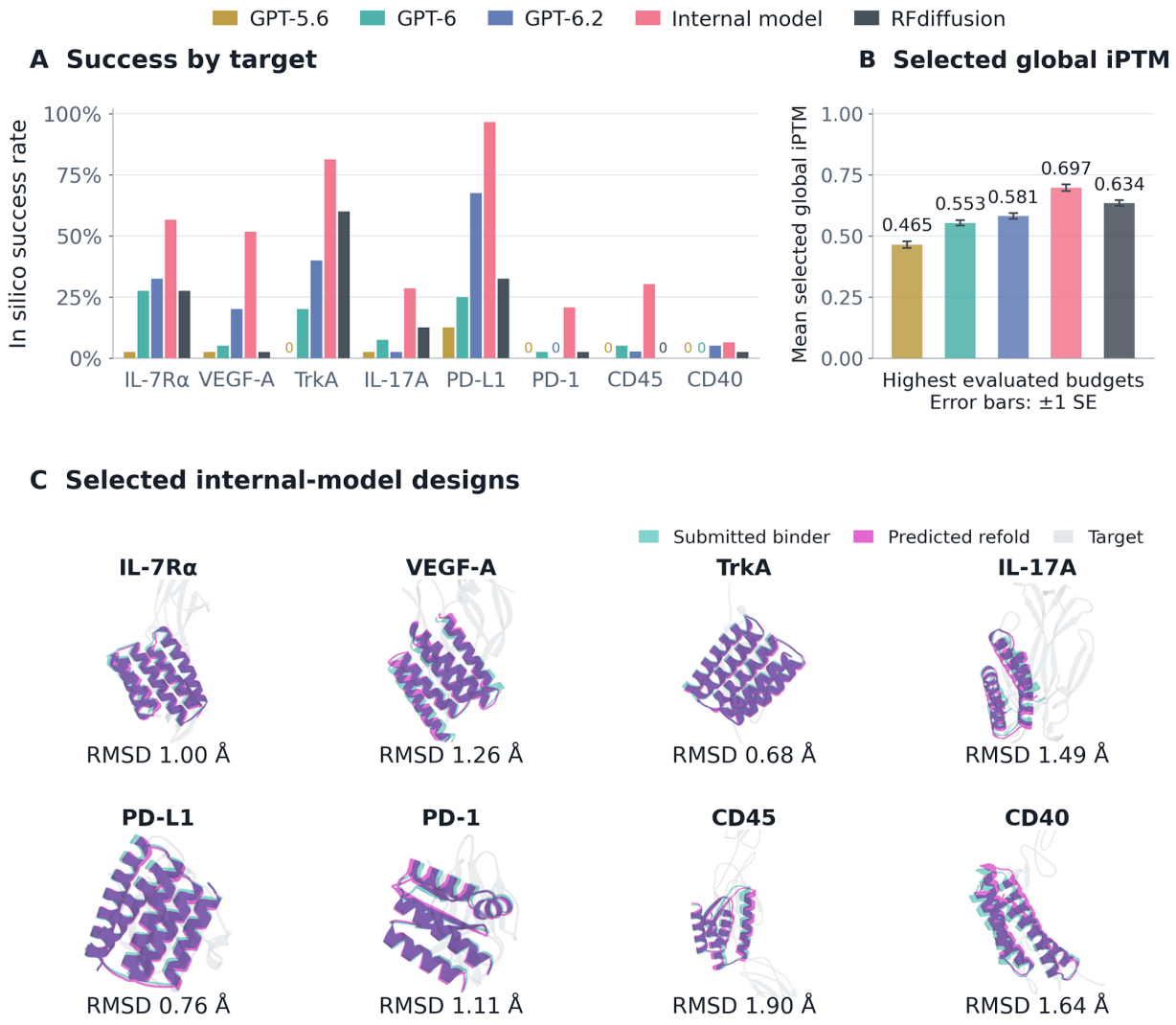


Figure 3. Minibinder design in silico. (A) Success rates by target at each model's highest evaluated budget. (B) Mean selected global ipTM; error bars show ±1 standard error. For each submitted backbone, eight SolubleMPNN sequence redesigns were refolded with ESMFold2 (Candido et al. 2026), and one was selected using a composite of global ipTM, complex pTM and binder-only backbone self-consistency (RMSD). Bars average the selected candidates' raw global ipTM across included trials pooled over eight targets. Colors identify the same models in A and B. (C) One internal-model design per target. Submitted backbones are overlaid with computationally predicted refolds.

2.1.2 Precise Control Over Binder Structures for Programmable Structure Design

In computational evaluations, binders designed by our models consistently contacted the specified target hotspots. Prompts could also specify which target residues to bury or request a β-sheet interface. We tested control over the binder fold itself.

We requested a structural class (mainly α, mainly β or α/β) or an exact count of three to eight

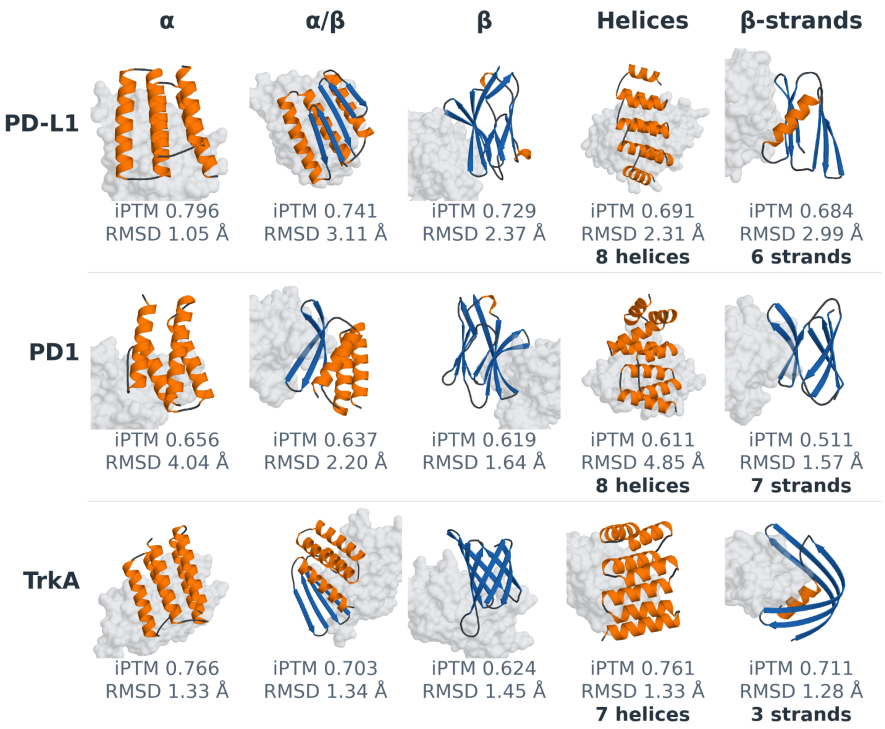


Figure 4. Representative refolded binders for PD-L1, PD-1, and TrkA are shown across three fold instructions and the highest qualifying helix and β-strand counts. Helices are orange, β-strands are blue, and targets are light gray. Each panel shows selected ipTM and target-aligned binder Cα RMSD; the displayed helix and strand counts match the requested counts.

α-helices or β-strands. Among designs that met the computational success criteria in Section 8.2, 100% matched the requested constraint. Representative qualifying designs are shown in Figure 4.

2.2 De Novo Minibinder Design Against Small-Molecule, RNA, and DNA Targets

Dataset and Settings. The task is to generate protein binder backbones targeting non-protein molecules, including RNA, DNA, and small molecules. We selected representative targets from previous studies, including RFdiffusion3 (Butcher, Krishna, et al. 2026) and BoltzDesign (Cho et al. 2025), covering small molecules, DNA, and RNA. We provided all-atom structures for nucleic acid targets, and SMILES only for small-molecule targets. As in Section 2.1, GPT had no access to the internet. As a representative specialized generative model, we compared GPT against RFdiffusion3.

Evaluations. We assessed the selected refold for each backbone using global ipTM and a measure of pose accuracy specific to the target type (Section 8.2).

2.2.1 Results

Our internal models match the performance of RFdiffusion3 across the evaluated non-protein targets (Figure 5). Although in silico success rates have improved considerably in our latest internal

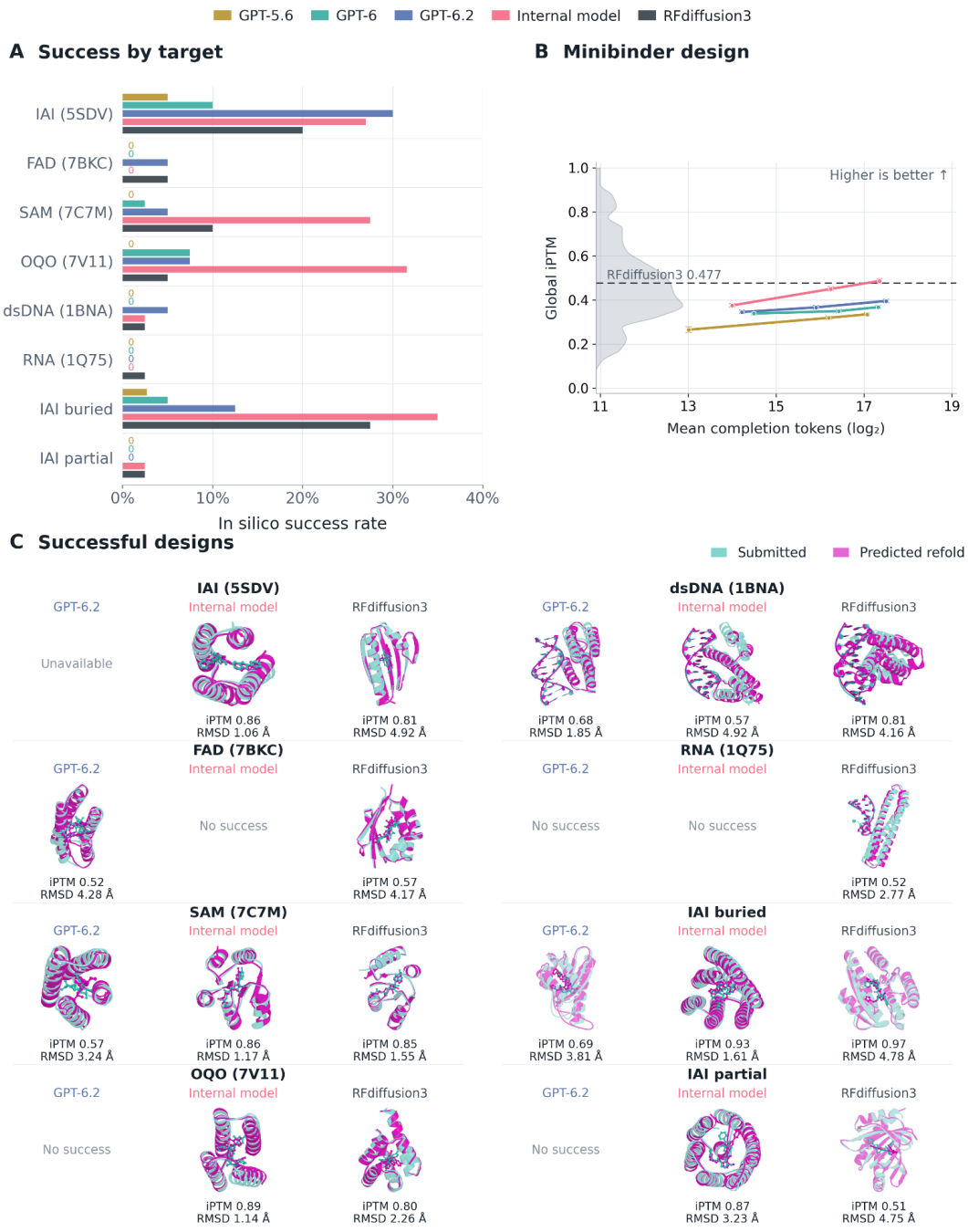


Figure 5. De novo minibinder design against small-molecule, RNA, and DNA targets. (A) In silico success rates across eight target settings for GPT models and RFdiffusion3. (B) Mean global ipTM versus $\log_2$(mean completion tokens), with error bars showing ± 1 standard error and the dashed line marking the RFdiffusion3 mean. (C) Representative computationally successful designs, with submitted structures (cyan) overlaid with ESMFold2-predicted refolds (magenta); labels report ipTM and modality-specific RMSD.

models, there is room for improvement. Figure 5 shows several computationally successful designs, illustrating GPT's promising intrinsic ability to design protein binders for diverse molecular

modalities.

2.3 Topology-Conditioned Backbone Generation

Dataset and Settings. We task the models with generating protein backbone structures conditioned on a CATH classification code and protein length. CATH is a hierarchical classification system for protein domains: Class (C), which describes the overall secondary-structure composition (e.g., mainly α-helical or β-sheet); Architecture (A), which describes the overall spatial arrangement of secondary structures; Topology (T), which describes the connectivity and arrangement of secondary-structure elements; and Homologous Superfamily (H), which groups domains inferred to share a common evolutionary ancestor (Orengo et al. 1997). We sampled the requested chain length at random and compared our models with Proteina (Geffner et al. 2025).

Evaluations. We evaluated the generated structures along two complementary dimensions: structural plausibility and agreement with the requested CATH classification. First, we assessed structural plausibility by measuring self-consistency RMSD (scRMSD) after direct backbone alignment (Section 8.2), following Campbell et al. (2024) and Watson et al. (2023). Second, we measured CATH classification agreement, which evaluates whether the generated structures match the requested structural classification at the class (C), architecture (A), and topology (T) levels. We excluded the homologous superfamily (H) level, as it may not be reliably inferred from backbone structures alone (Orengo et al. 1997). We used Merizo (Lau et al. 2023) to identify structural domains in the refolded proteins. Finally, we defined the success rate as the fraction of generated structures with an scRMSD below 2.0 Å for which Merizo-Search (Kandathil et al. 2025) identified the correct CATH topology in the refolded backbone.

2.3.1 Results

At the highest setting tested, predicted refolds often closely reproduced the backbones generated by our internal model (median scRMSD 1.65 Å, Figure 6B). Some of these refolds matched the requested CATH class and architecture but differed in topology. In 18.9% of designs, at least one refold matched the requested topology and had an scRMSD below 2 Å, compared with 17.1% for Proteina (Figure 6C).

2.3.2 Predicting CATH Classifications from Backbone Coordinates

While the previous results demonstrate that GPT models are not yet perfect at generating structures conditioned on CATH classifications, a natural question is whether they can solve the reverse task: predicting the CATH classification of a given protein backbone structure. This reverse task is particularly interesting because it tests whether models can recognize and reason about global protein structural features. It also provides a way to assess whether models can serve as structural verifiers for topology-conditioned backbone generation.

Dataset and Settings. We evaluated the reverse task by providing protein backbone coordinates and asking models to predict the corresponding CATH classifications using protein structures from RCSB PDB (Berman et al. 2000).

Evaluations. We evaluated prediction accuracy across the CATH hierarchy, primarily focusing on

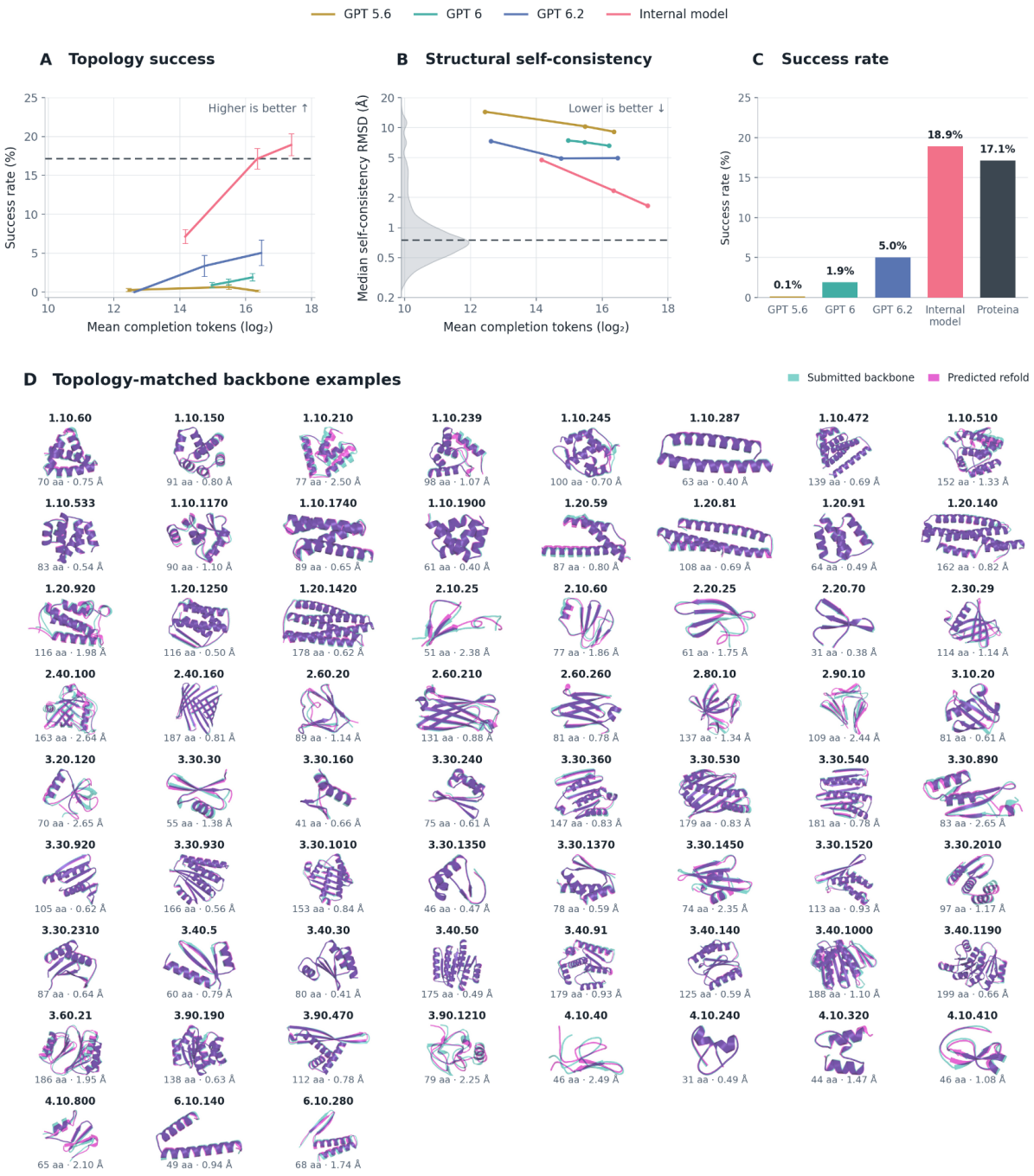


Figure 6. Topology-conditioned backbone generation: (A, B) topology success and median backbone self-consistency RMSD versus $\log_2$(mean completion tokens), and (C) success rate at the highest evaluated settings, with Proteina as a reference. Success requires the requested CATH topology and RMSD < 2 Å for the same candidate among up to eight refolds. (D) Submitted backbones (cyan) and predicted refolds (magenta) are overlaid for topology-matched examples.

four levels: class, architecture, topology and superfamily. We defined a hierarchical agreement score as follows:

- 1: Correct homologous superfamily (H-level agreement).
- 3/4: Correct topology (T-level agreement), but incorrect superfamily.
- 2/4: Correct architecture (A-level agreement), but incorrect topology.
- 1/4: Correct class (C-level agreement), but incorrect architecture.
- 0: Incorrect class.

Note that H-level agreement may be difficult to establish from backbone structures alone, as homologous superfamily assignment often requires additional sequence or evolutionary information.

Results.

Interestingly, as shown in Figure 7, while earlier models struggled substantially with this task, recent models demonstrate much stronger capabilities in recognizing structural classifications directly from backbone coordinates. In particular, recent models can predict not only broad structural classes and architectures but also, in some cases, the more fine-grained topology of protein domains like homologous superfamily. These findings suggest that recent models have developed a meaningful understanding of global structural organization in proteins, even though translating this understanding into accurate coordinate generation remains challenging.

2.4 Motif Scaffolding

Dataset and Settings. The task is to generate protein scaffold backbones around predefined structural motifs while preserving the geometry of those motifs. Motif scaffolding is important in enzyme design, where catalytic residues must be positioned precisely in three-dimensional space, and in binder design, where specific binding motifs must be maintained within a stable protein scaffold (Watson et al. 2023).

We used MotifBench (Zheng et al. 2025), a benchmark for evaluating motif-scaffolding methods, to define the structural motifs and evaluate generated scaffolds. Depending on the task, motifs may be specified using backbone coordinates alone or include all-atom coordinates, including side chains. The models were tasked with generating the remaining scaffold backbone while preserving the specified motif geometry.

We compared GPT with RFdiffusion (Watson et al. 2023).

Evaluations. A scaffold passed if at least one of its eight refolds met both the motif RMSD and scaffold self-consistency RMSD (scRMSD) cutoffs (Section 8.2).

2.4.1 Results

The internal model achieved a higher pooled success rate than RFdiffusion across evaluable scaffolds (43.7%, 296/677, versus 27.8%, 790/2,845). Its success rate was higher on 19 of 30 motifs, lower on eight and tied on three (Figure 8).

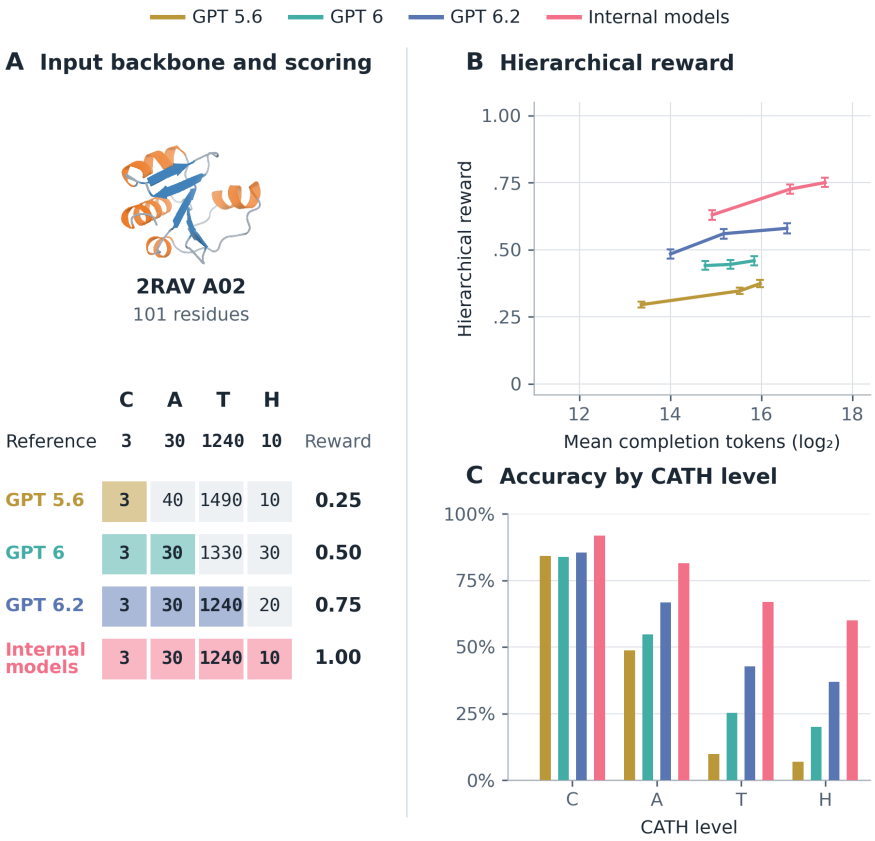


Figure 7. Protein backbone classification across the CATH hierarchy. (A) A selected example (2RAV A02; Lu et al. 2008) illustrates CATH assignments from backbone coordinates. Colored cells mark the correct prefix. The score increases by $\frac{1}{4}$ for each consecutive correct level from class (C) through architecture (A), topology (T), and homologous superfamily (H). Orange and blue show helices and strands. (B) Mean hierarchical reward versus $\log_2$(mean completion tokens), with error bars showing ± 1 standard error across 399–400 domains. (C) Cumulative accuracy on the same domains at each model's highest evaluated setting; settings are not compute matched. C/A/T/H accuracies (%) are GPT-5.6: 84.25, 48.75, 9.75, 7.00; GPT-6: 83.75, 54.75, 25.25, 20.00; GPT-6.2: 85.50, 66.75, 42.75, 37.00; Internal models: 91.75, 81.50, 67.00, 60.00.

2.5 Wet-Lab Studies

We tested backbone designs for five protein targets.

- Backbone design: Given the target structures, we used a pre-release version of GPT-6.2 to design candidate backbones for target–binder complexes, following the procedure described in Section 2.1.
- Sequence design: We used SolubleMPNN (Goverde et al. 2024) to generate eight candidate binder sequences for each backbone.
- We selected designs within each target for experimental testing (Section 8.3).

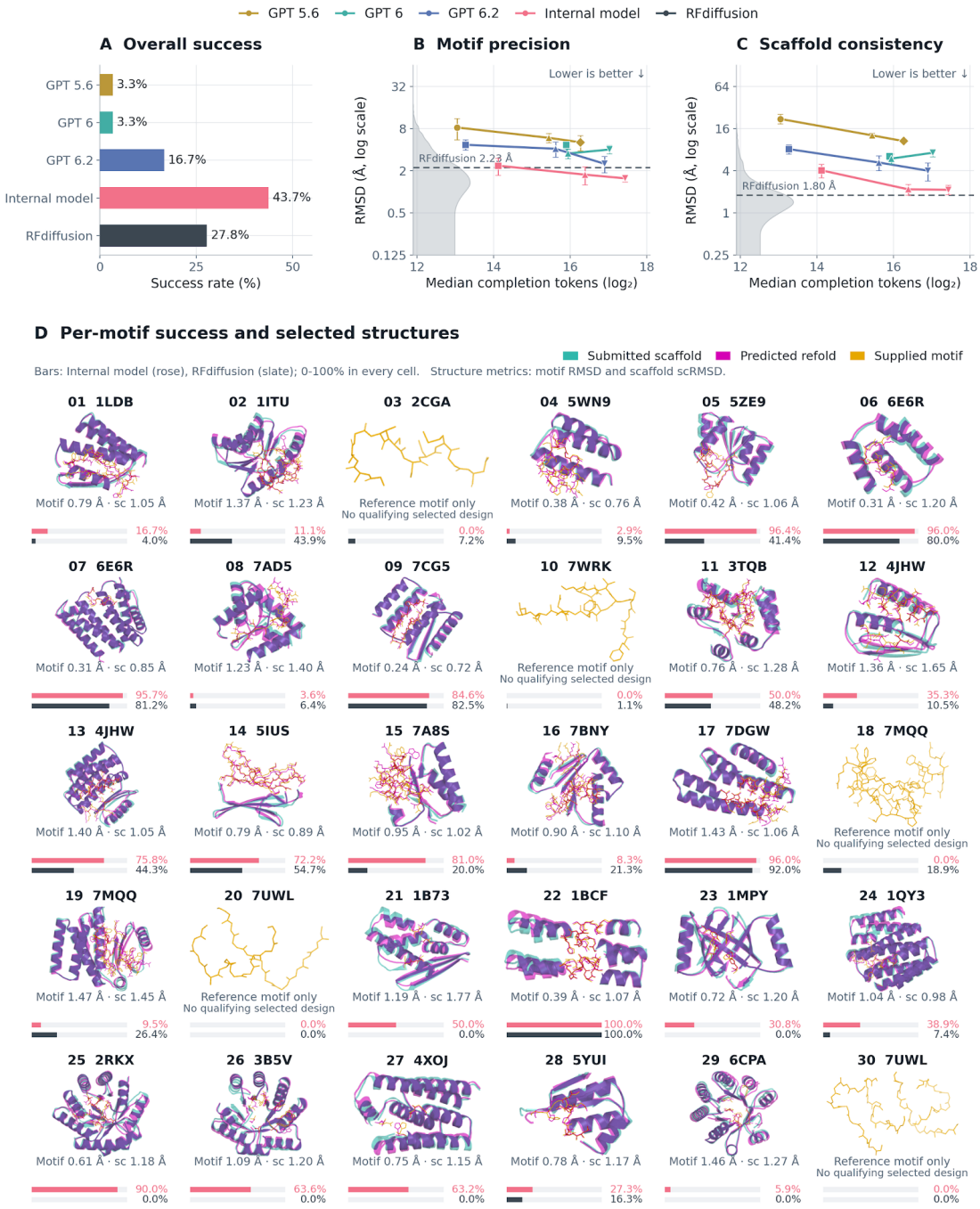


Figure 8. Motif scaffolding on MotifBench. (A–C) Overall success, motif RMSD, and scaffold self-consistency RMSD (scRMSD) compare GPT models with RFdiffusion; success requires motif RMSD < 1.5 Å and scRMSD < 2 Å. (D) Per-motif success rates and selected structures are shown with submitted scaffolds in teal, predicted refolds in magenta, and supplied motifs in orange; five cases show only the reference motif because no selected design qualified.

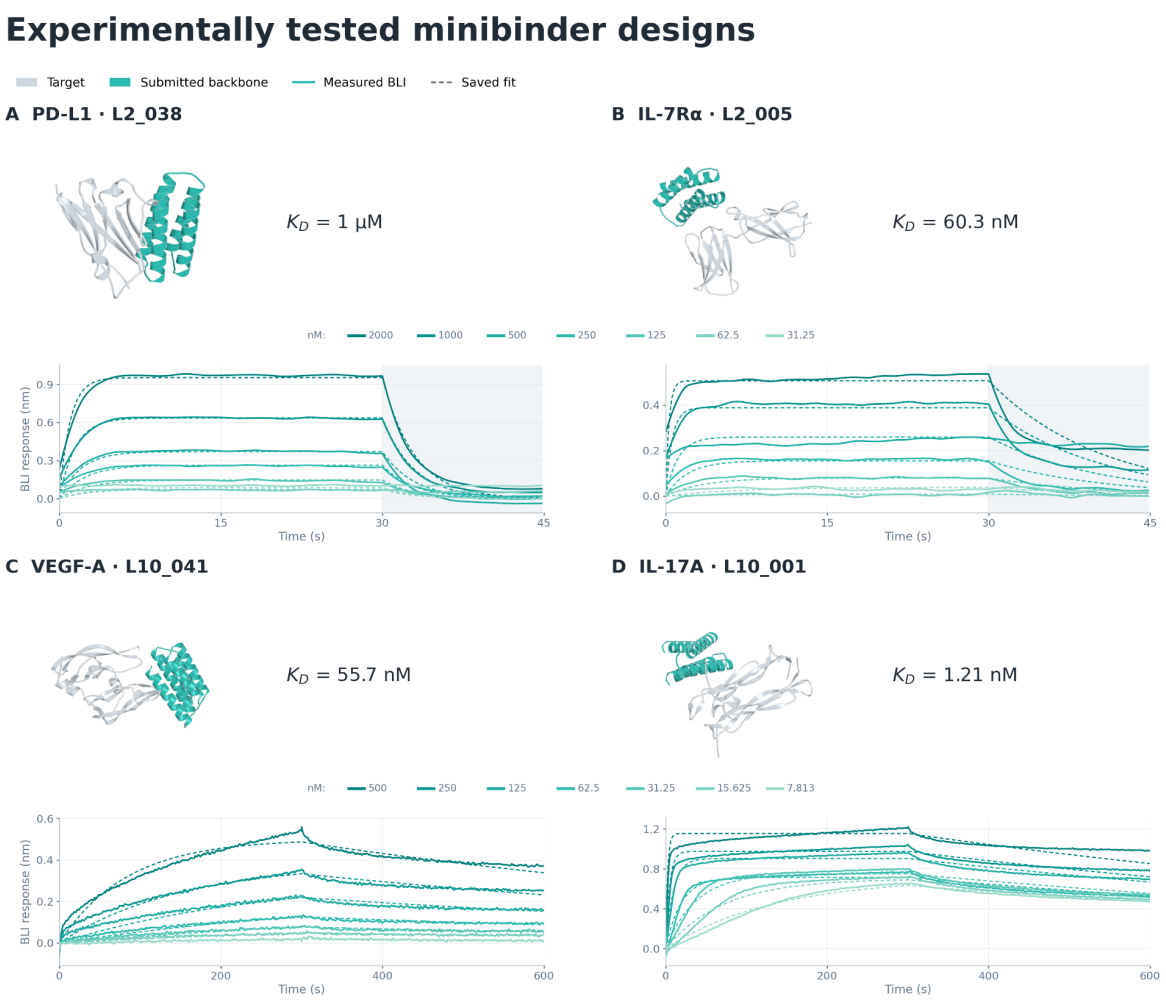


Figure 9. Experimentally tested minibinder designs. Submitted binder backbones (teal) and targets (light gray) are shown for (A) PD-L1 L2_038, (B) IL-7Rα L2_005, (C) VEGF-A L10_041, and (D) IL-17A L10_001. Solid lines show measured BLI responses and dashed lines show fitted curves at the indicated concentrations, with reported KD values of 1 μM, 60.3 nM, 55.7 nM, and 1.21 nM, respectively.

Wet-Lab Results:

In the 1 μM biolayer interferometry (BLI) screen, the hit rates were 5/92 (5.4%) for PD-L1, 3/93 (3.2%) for IL-7Rα, 1/46 (2.2%) for TrkA, 8/46 (17.4%) for VEGF-A, and 18/45 (40.0%) for IL-17A. Follow-up measurements using multi-concentration BLI yielded a dissociation constant of 1.21 nM for the IL-17A binder shown in Figure 9. For details on the experimental procedures, see Appendix B.

3 Capability Analysis: Protein Sequence Design

We start by analyzing inverse folding capabilities. Inverse folding models have been widely used to generate functional protein variants in protein design, including affinity maturation in antibody design (Shanker et al. 2024). Surprisingly, GPT-6.2 and our internal models demonstrate strong inverse folding capabilities, matching state-of-the-art specialized models on our benchmarks.

3.1 Protein Inverse Folding

Dataset and Settings. The task here is to generate protein sequences given backbone coordinates (N, CA, C, O). We evaluated GPT-5.6, GPT-6, GPT-6.2, and our internal models in the standard setting. We compared our models against a state-of-the-art specialized model, ProteinMPNN (Dauparas et al. 2022).

We used subsets of two datasets: RCSB PDB (Berman et al. 2000), which provides experimentally determined backbone structures, and AFDB (Varadi et al. 2022), which provides AlphaFold-predicted backbone structures. Note that RCSB PDB includes both monomeric and multimeric protein structures.

Evaluations. We measured native sequence recovery and structural self-consistency after refolding. We report TM-score (Zhang and Skolnick 2004) and the fraction of evaluable targets for which every required chain had a backbone RMSD below 2 Å (Section 8.2).

3.1.1 Results

GPT-6.2 and our internal models achieve strong sequence recovery rates on both RCSB PDB and AFDB, generally outperforming ProteinMPNN (Figure 10A–B). Their self-consistent TM-scores and success rates are comparable to those of ProteinMPNN. The combination of slightly worse self-consistency and high recovery rate suggests potential bias due to memorization rather than generating truly compatible sequences (Figure 10C–D).

We make a few observations regarding memorization on this task. First, many RCSB data points have sequence recovery rates close to 1.0, suggesting memorization. However, this memorization is not shallow, as the models still need to interpret input backbone structures to recover the original sequences by reasoning over backbone structures, motifs, and physical principles—a highly nontrivial task even for expert structural biologists. Consistent with this difficulty, GPT-5.6 exhibits relatively low recovery rates.

We further hypothesize that the higher sequence recovery rates in AFDB are due to our internal model's ability to infer residue identities from geometric patterns introduced during structural relaxation. AFDB contains predicted structures from AlphaFold, followed by a relaxation step. Analysis of model reasoning traces suggests that this model consistently recognized relaxation artifacts to recover nearly 95–100% of the original sequences.

The extent to which our models can generalize beyond memorized structures must ultimately be addressed by evaluating them on unseen, non-public structures that have low structural similarity to known proteins. Nevertheless, we highlight four characteristic behaviors that suggest the models have developed a meaningful understanding of protein structures: (1) the models do not

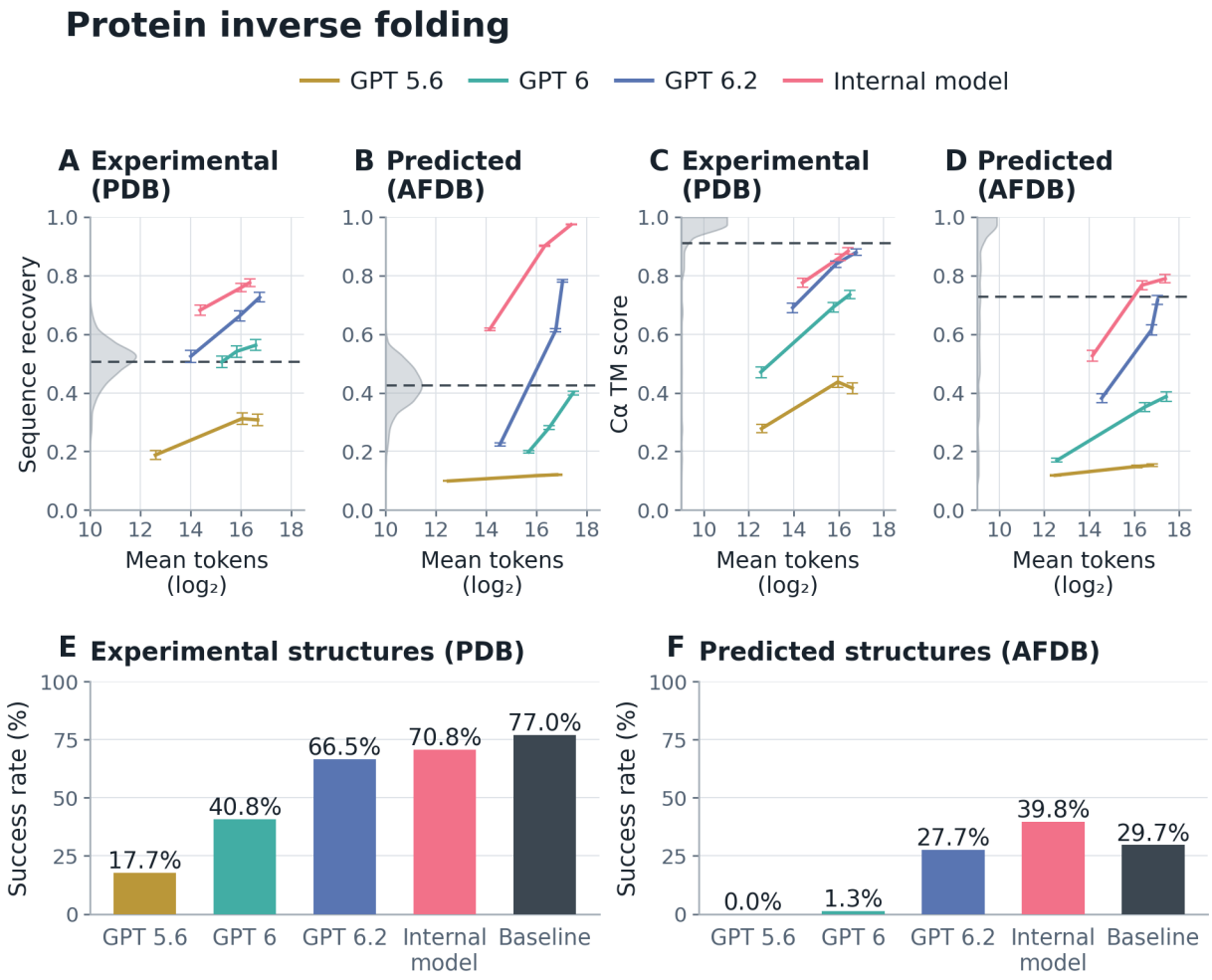


Figure 10. Protein inverse folding on experimentally determined RCSB PDB backbones (A, C, E) and AlphaFold-predicted backbones from the AlphaFold Database (AFDB; B, D, F). The PDB cohort includes monomers and multimers. A, B: Native sequence recovery. C, D: Refolded Cα TM-score. Curves show means; error bars denote ± 1 standard error. The horizontal axis in A–D is $\log_2$(mean completion-token count). Gray profiles and dashed lines show ProteinMPNN distributions and means: 0.506 (A), 0.425 (B), 0.910 (C), and 0.728 (D). E, F: Success rates at each language model's highest evaluated setting; baseline bars show ProteinMPNN.

simply memorize 3D coordinates, (2) they respond appropriately to biologically meaningful molecular contexts, (3) they perform well on non-publicly available synthetic structures, and (4) they demonstrate strong protein redesign capabilities. In this section, we focus on points (1), (2), and (4), while the remaining points are addressed in Section 3.3.

3.1.2 GPT Models Understand SE(3) Transformations

Models with meaningful understanding of protein structures should produce predictions invariant to SE(3) transformations (rotations and translations of the input coordinates). Indeed, we observe that sequence recovery rates remain nearly unchanged after SE(3) transformations (Figure 11).

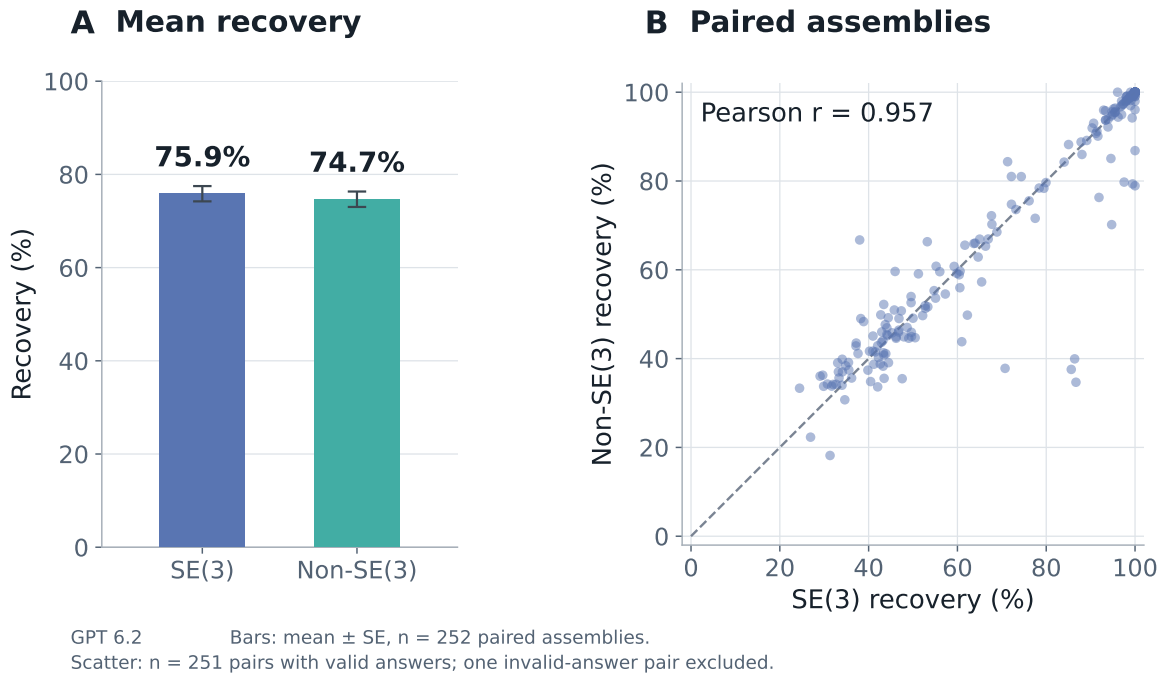


Figure 11. Inverse-folding sequence recovery under rigid-body transformations of the input coordinates. (A) GPT-6.2 shows similar mean recovery with and without SE(3) rotations and translations (75.9% versus 74.7%; mean ± standard error, 252 paired assemblies). (B) Recovery rates for the 251 pairs with valid answers are strongly correlated (Pearson r = 0.957); the dashed line indicates equal recovery.

Furthermore, the Pearson correlation between recovery rates before and after transformation is very high across a subset of evaluated samples. These results suggest that the models are not simply memorizing absolute 3D coordinates, but instead can reason about the 3D geometry of proteins.

Importantly, unlike specialized neural-network architectures such as ProteinMPNN, where SE(3) invariance is explicitly built into the architecture (Dauparas et al. 2022), our models naturally exhibit this behavior through reasoning, without requiring explicit architectural constraints.

3.1.3 Models Take Molecular Contexts Into Account

When performing inverse folding, GPT models can incorporate diverse molecular contexts, including ligands, DNA, RNA, and other proteins, similar to LigandMPNN (Dauparas et al. 2025). To investigate this capability, we conducted experiments in two settings: protein–protein complexes (heterodimers) and protein–ligand complexes. In each setting, we compared two conditions: with molecular context, where the model receives the target protein backbone along with interacting proteins or ligands, and without molecular context, where only the target protein backbone is provided.

Providing the binding partner increased mean interface sequence recovery from 75.7% to 80.6% for protein partners and from 68.7% to 76.0% across small molecules, DNA, and RNA combined (Figure 12). The latter finding is consistent with the higher recovery reported for LigandMPNN

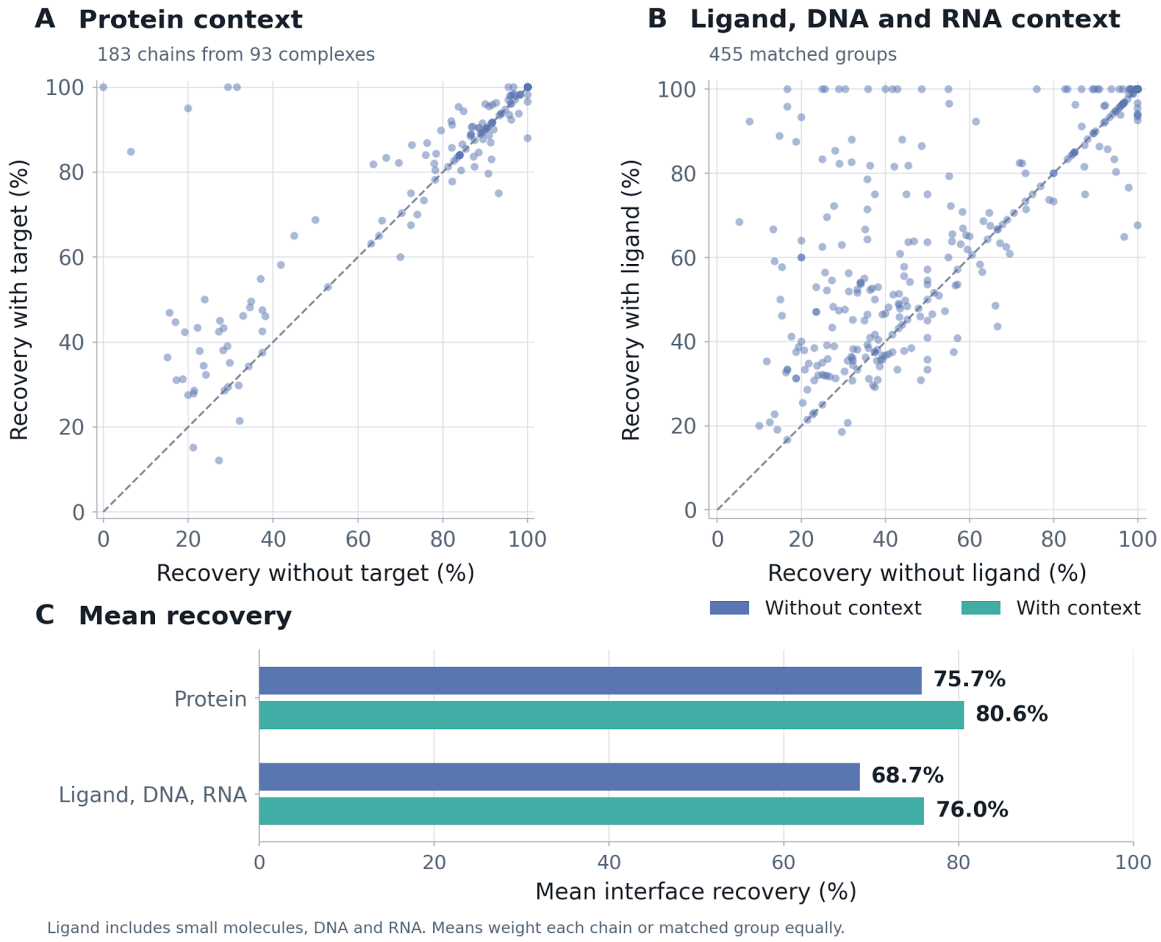


Figure 12. Interface sequence recovery with and without molecular context. Paired results are shown for protein partners (A; 183 chains from 93 complexes) and small molecules, DNA, or RNA (B; 455 matched groups), with dashed diagonals indicating equal recovery. (C) Mean recovery increases from 75.7% to 80.6% for protein interfaces and from 68.7% to 76.0% for the combined ligand, DNA, and RNA set; each chain or matched group is weighted equally.

than for ProteinMPNN at small-molecule and nucleic-acid binding sites (Dauparas et al. 2025).

3.1.4 GPT Can Redesign Sequences

To further investigate generalization beyond memorization in intrinsic inverse folding, we provided our models with paired backbone structure and wild-type sequences from RCSB PDB and requested redesigned sequences that remain compatible with the given structure while satisfying mutation ranges (25–45%, 45–55%, 55–65%, 65–75%, and 75–85%). For evaluation, we again used ESMFold2 to compute self-consistency RMSD, following the procedure described in Section 3.1. This metric can be nonzero even for the unmodified wild-type sequence because the Cα coordinates predicted by ESMFold2 can differ from those of the experimental backbone after alignment.

When asked to change 75–85% of residues, the models changed 75.00–75.65%, with a median self-

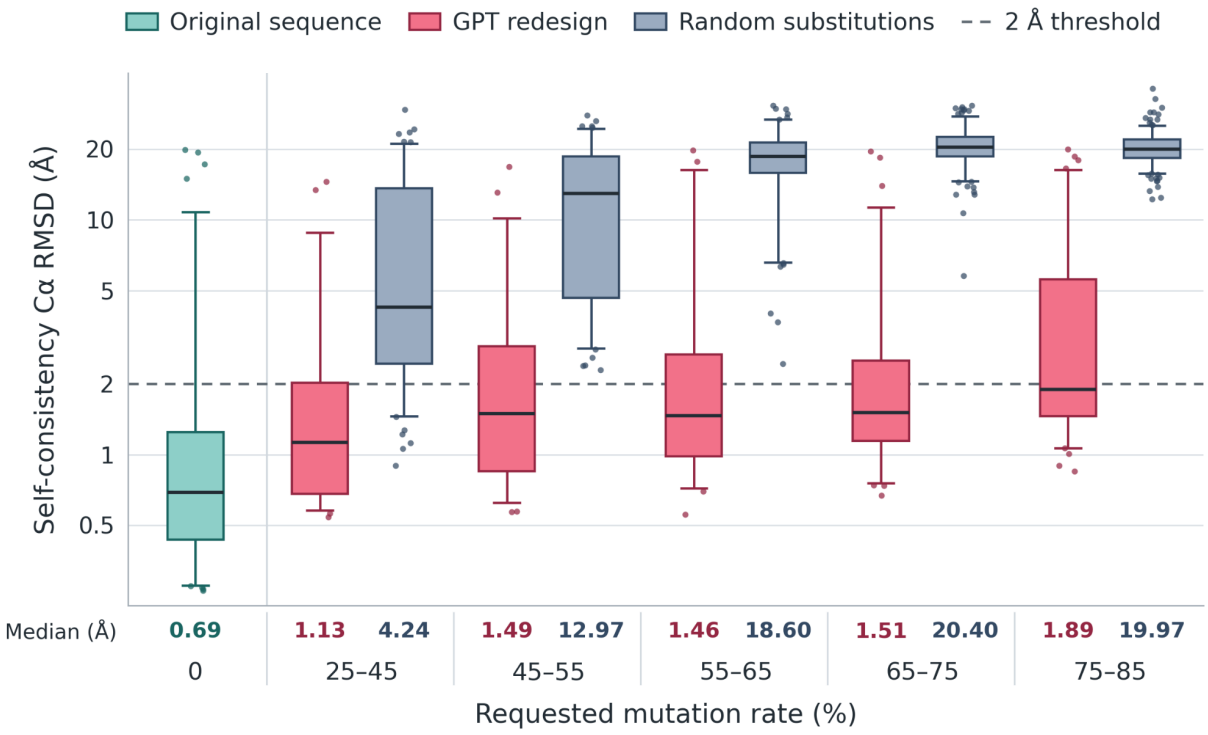


Figure 13. Structural self-consistency after sequence redesign at prescribed mutation rates. All 236 scored GPT designs on 70 backbones are compared with three random variants per design, matching its original sequence and actual substitution count. Mutation positions and alternative residues were sampled uniformly. Controls for the original sequences compare ESMFold2 predictions of the unmodified PDB sequences with the experimental backbones. Cα RMSD compares the aligned ESMFold2 prediction with the supplied backbone (log scale). Boxes show medians and quartiles, whiskers mark the 5th and 95th percentiles, and dots show more extreme values. Success requires RMSD < 2 Å (dashed line).

consistency Cα RMSD of 1.89 Å across all 76 scored designs in this range (Figure 13).

We compared all 236 scored GPT designs on 70 backbones with three random variants of the corresponding input sequence per design. Each variant matched the design's actual substitution count, with positions sampled uniformly without replacement and replacement residues sampled uniformly from the other 19 amino acids. The median RMSD for the unmodified input sequences was 0.69 Å across all 70 backbones. GPT designs had lower median RMSD than matched random variants in all five ranges. In the 75–85% range, the medians were 1.89 Å for GPT designs and 19.97 Å for random variants. RMSD was below 2 Å for 39 of 76 GPT designs and none of the 228 random variants.

Simple recall of training examples is unlikely to explain this performance.

This task highlights the flexibility of inverse folding with models that understand natural language. Performing this task with conventional MPNN-based methods is less straightforward, as it requires an additional strategy for determining which residue positions to mutate while satisfying the prescribed mutation rate. In contrast, our models can directly interpret these constraints and reason about which residues to mutate and which to preserve to maintain structural compatibility.

3.2 Antibody Inverse Folding

Dataset and Settings. The task is to generate CDR sequences given CDR backbone coordinates, framework backbone coordinates, and framework sequences. We used a dataset derived from SAbDab (originally sourced from PDB; Dunbar et al. 2014). We compared our models against AbMPNN (Dreyer et al. 2023), a representative specialized model for antibody inverse folding.

Evaluations. We measured CDR sequence recovery and structural self-consistency. Success required a mean RMSD below 2 Å across the applicable CDRs of each antibody (Section 8.2).

3.2.1 Results

GPT-6.2 and our internal models achieve higher sequence recovery than AbMPNN (Figure 14B) with comparable mean self-consistency RMSD and success rates, mirroring the results for general proteins (Section 3.1). This pattern holds across CDRs, including HCDR3, the most structurally diverse antibody CDR (Tsuchiya and Mizuguchi 2016).

3.3 Wet-Lab Studies

We tested minibinder inverse folding, antibody inverse folding, and therapeutic antibody redesign. Section 8.3 describes how we selected designs for testing.

3.3.1 Minibinder Inverse Folding

Methods. We asked a pre-release version of GPT-6.2 to design sequences for experimentally confirmed minibinders from nonpublic target–binder backbones, withholding the parental binder sequences. For PD-L1 and IL-7Rα, we used backbones generated by pre-release GPT-6.2 (Section 2.5). For PD-1, VEGF-A, and MDM2, we used ESMFold2 to predict complex structures from binders previously identified by Onava.

Across the nine binder families that yielded experimental hits, ESMFold2 success rates ranged from 1.1% to 79.3% of evaluable designs (Figure 15). Many designs achieved high ipSAE scores while retaining half or fewer of the parental residues.

Sequence diversity varied even among families with similar success rates. Both MDM2 families had ESMFold2 success rates above 70%, yet passing sequences differed from one another at an average of 40% of positions in one family and only 4% in the other. The less diverse family also had higher median parental sequence recovery (91% versus 55%). By comparison, mean pairwise differences ranged from 43% to 48% across the three VEGF-A families.

Wet-Lab Results

Screening yielded 123 hits across five targets. With GPT-generated backbones, hit rates were 10.1% for PD-L1 (9/89) and 27.8% for IL-7Rα (5/18). With ESMFold2-predicted backbones, hit rates were 13.6% for PD-1 (8/59), 61.9% for VEGF-A (26/42), and 85.2% for MDM2 (75/88).

Some designs bound more tightly than their parents. For example, one VEGF-A design had a fitted

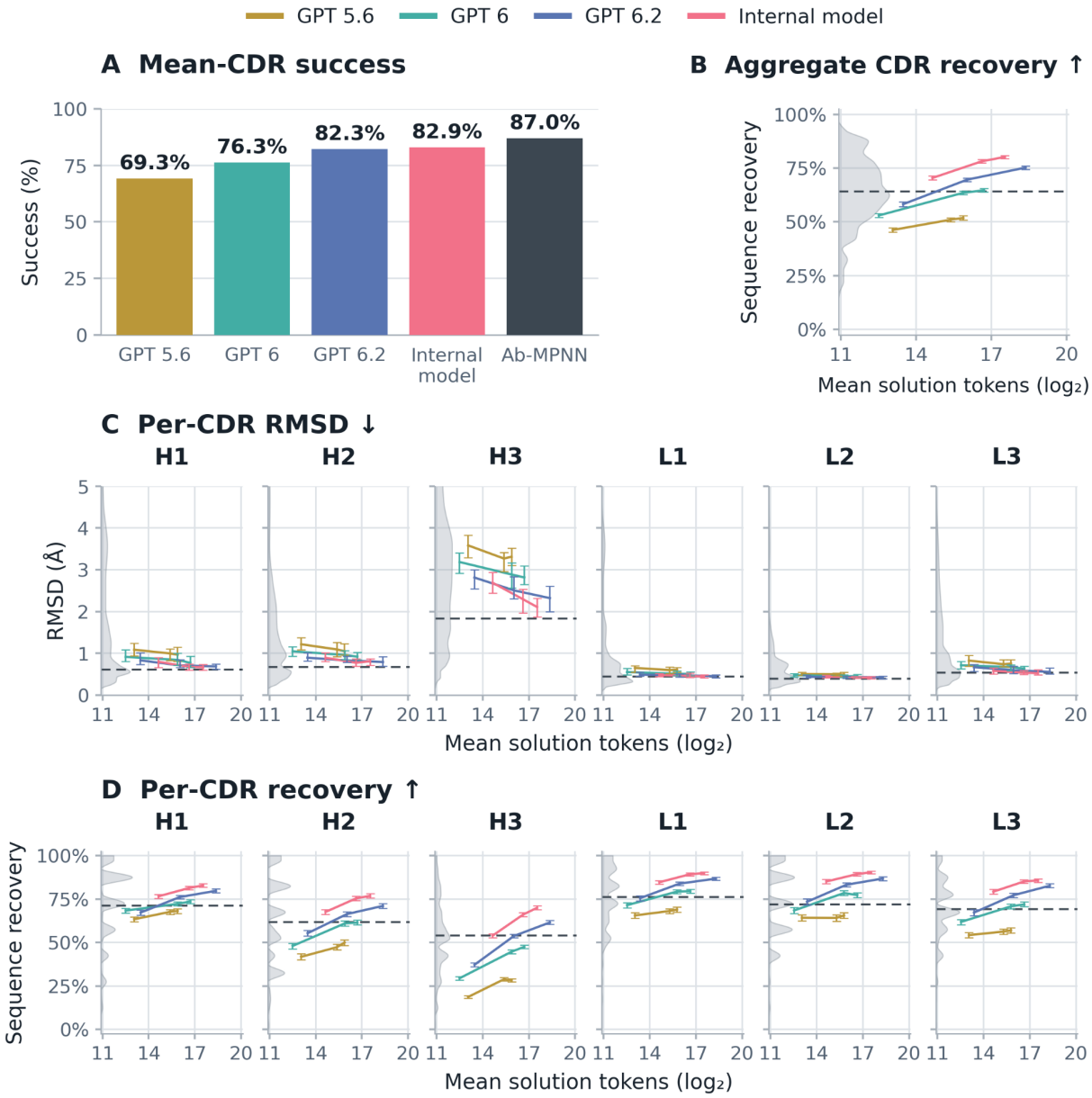


Figure 14. Antibody inverse folding using structures from SAbDab. (A) Success requires mean self-consistency RMSD across applicable CDRs < 2 Å for each antibody, excluding incomplete tasks; (B) aggregate sequence recovery and (C, D) per-CDR RMSD and recovery are compared with AbMPNN. Recovery points show means ± standard error, RMSD points show medians with 95% confidence intervals, and gray densities and dashed lines summarize the AbMPNN baseline.

K_D of 0.21 nM, compared with 3.13 nM for its parent in the same replicate (Figure 16). Of the 123 screening hits, 38 shared only 30–50% sequence identity with their parental binders (Figure 17). These hits spanned four of the five targets and both backbone sources.

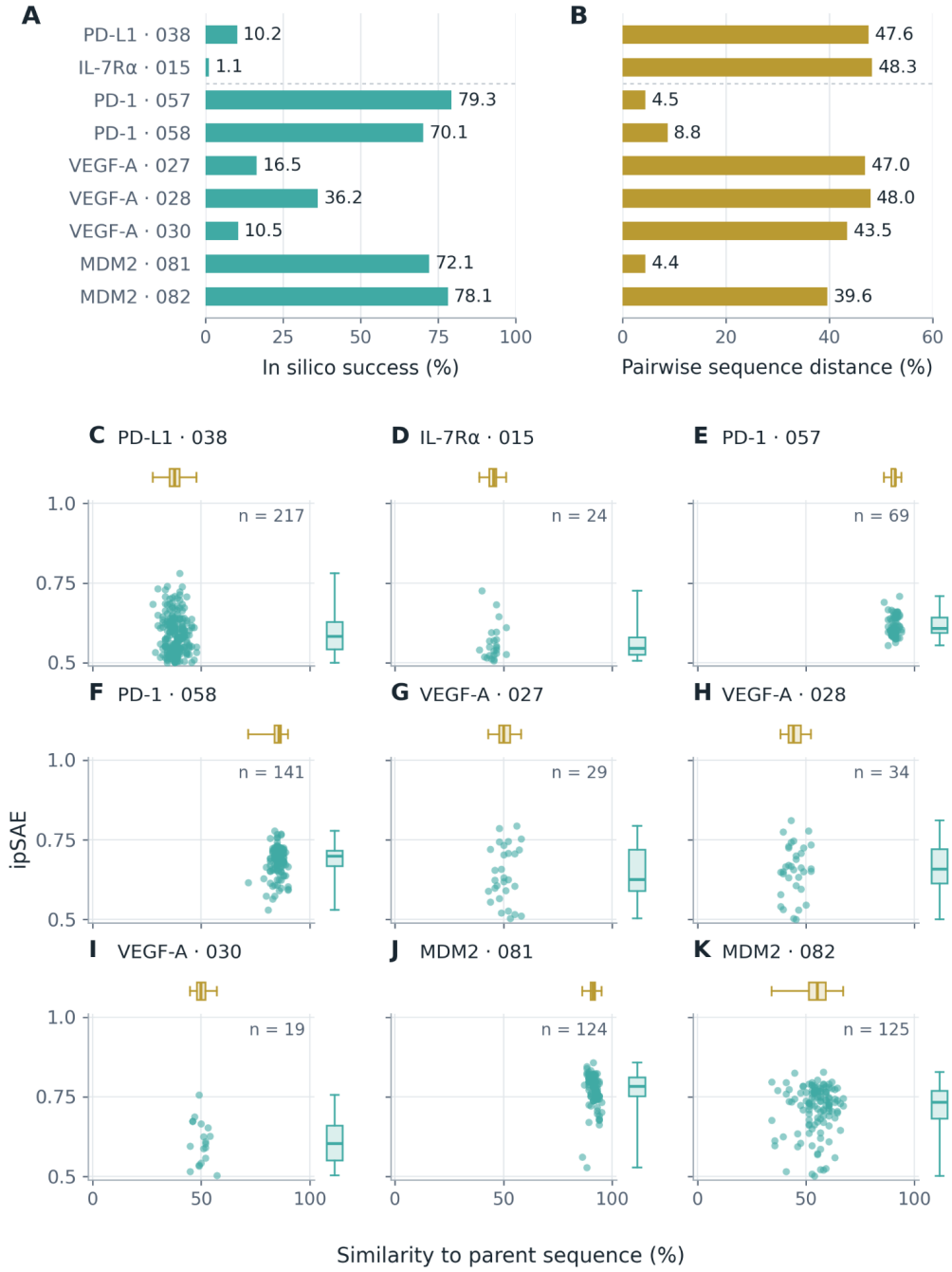


Figure 15. In silico success and sequence diversity across nine minibinder families. (A) ESMFold2 success rates among 5,439 evaluable designs. (B) Mean pairwise sequence distance among passing designs. (C–K) ipSAE (Dunbrack 2025) versus similarity to the parent sequence (percent positional amino-acid identity) for 782 passing designs; each point represents one record, including repeated sequences, and overlapping points are retained. Marginal boxes show medians and interquartile ranges; whiskers span observed minima to maxima. PD-L1_038 and IL-7Rα_015 use backbones generated by pre-release GPT-6.2; others uses ESMFold2-predicted complexes.



Figure 16. BLI measurements for representative inverse folding designs against VEGF-A and MDM2, shown above their parental binders. Colors indicate antigen concentration, solid lines show measured responses, and dashed lines show fitted curves, with KD values labeled.

3.3.2 Inverse Folding Capabilities in Antibody Design

We used nonpublic backbones predicted for complexes of IL-4Rα with VHHs previously identified by Onava (Appendix C).

- Sequence generation. For each reference antibody–antigen complex, GPT-6.2 designed sequences for all three HCDRs using the antigen sequence and the antibody framework structure and sequence. Parental HCDR sequences were withheld.



Figure 17. Parental sequence recovery across nine binder families, shown in bins spanning 5 percentage points for all 233 designs sent for testing (gray) and 123 binder calls (teal). Recovery is the percentage of residues matching the parental sequence; dashed and dotted lines mark the respective medians. Four sent designs lack reported results; binder calls include atypical screening hits.

- We selected candidates for testing as described in Section 8.3.

Across seven parental VHH groups, 9,902 designs passed the ESMFold2 screen. Many had high ipSAE scores despite low sequence identity to the parental CDRs (Figure 18).

Wet-Lab Results

Of the 184 VHH designs submitted for testing, 182 had a reported result in the single-dose screen. Of these, 128 (70.3%) screened positive for IL-4Rα binding. We retested six screening hits by



Figure 18. In silico success and CDR sequence diversity across seven parental VHH groups. **(A) Percentage of evaluable designs passing the ESMFold2 screen (ipSAE $\geq$ 0.5, self-consistency RMSD $\leq$ 2.0 Å for each CDR, and target-aligned binder RMSD $\leq$ 4.0 Å). (B) Mean pairwise CDR sequence distance among passing designs. (C–I) ipSAE versus parental CDR sequence recovery for the 9,902 passing designs. Counts indicate passing designs per group; marginal boxplots summarize the distributions. Recovery is the percentage of CDR residues matching the parent.**

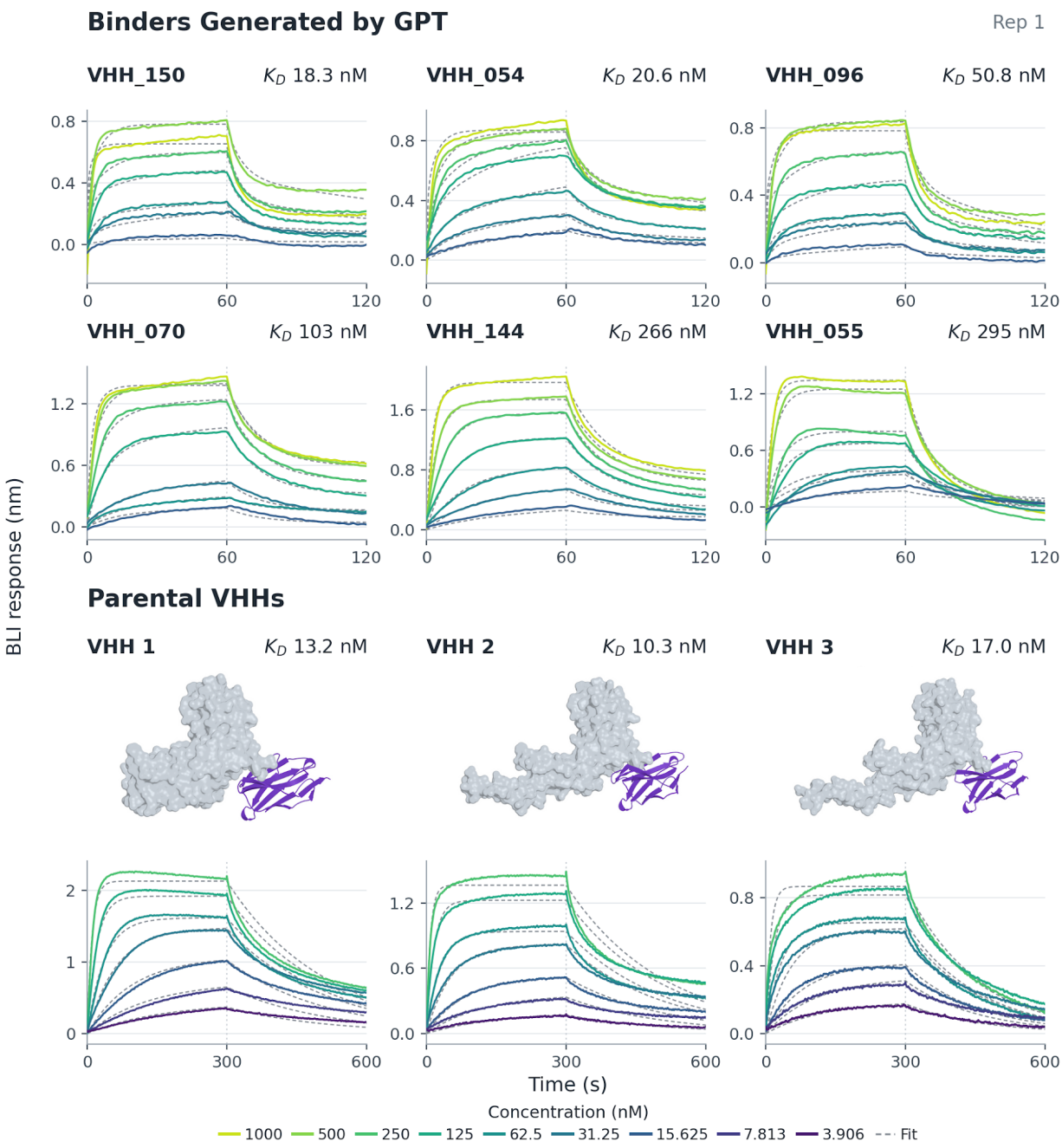


Figure 19. Biolayer interferometry sensorgrams for six confirmed IL-4Rα binders (top two rows) and three top parental VHHs with predicted complexes used for the campaign (bottom row). Colored lines, gray dashed fits, and K_D values are from replicate 1. Upper panels use the updated protocol with shorter association and dissociation; columns are not matched parent–design pairs.

dose–response biolayer interferometry; six confirmed binders are shown in Figure 19. Two variants of the same parental VHH, VHH 054 and VHH 150, each differed from the parent at 17 of 25 CDR positions. Both retained IL-4Rα binding, with apparent K_D values ranging from 10.3 to 21.8 nM across the two variants and two replicates.

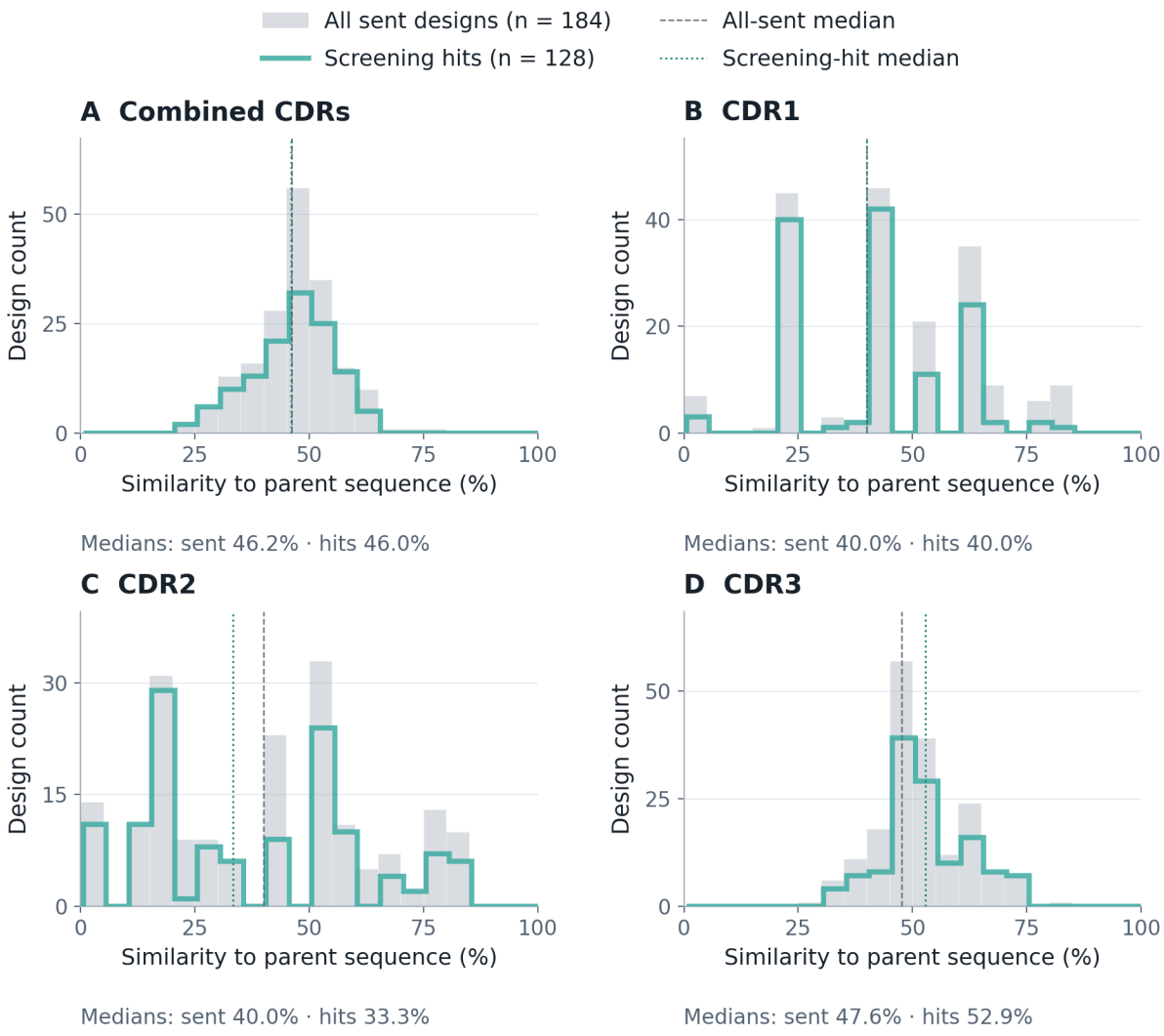


Figure 20. Similarity to parent sequence for all 184 sent designs (gray fill) and 128 screening hits (teal outline), across combined CDRs and CDR1–3. Similarity is percent identity. Histograms use 5-percentage-point bins; outlines are offset right by 0.6 percentage points for visibility. Dashed and dotted lines mark unshifted medians.

Among the 128 IL-4Rα screening hits, median sequence identity to the parental VHH was 46.0% across the three CDRs combined and 52.9% in CDR3 (Figure 20). Appendix D.1 shows variation at each CDR position across the nine VHH families (Figure A3A–C).

3.3.3 Redesigning Capabilities for Therapeutic Antibodies

We next evaluated antibody redesign in the lab, extending the in silico task in Section 3.1.4 to six therapeutic antibodies targeting PD-1 or PD-L1. We used reference complexes from two sources.

- Experimental structures from the RCSB PDB. Pembrolizumab–PD-1 (5JXE; Na et al. 2017), tislelizumab–PD-1 (7CGW; Hong et al. 2021), serplulimab–PD-1 (7E9B; Issafras et al. 2021),

and avelumab–PD-L1 (5GRJ; Liu et al. 2017).

- Predicted structures. We used folding models (ESMFold2) to generate seed backbones for finotonlimab–PD-1 and cosibelimab–PD-L1. Neither complex has an experimentally determined structure deposited in the PDB, so these cases required structure prediction followed by sequence redesign.

We used the same two-step procedure for experimental and predicted complexes.

- Sequence generation. We supplied the reference complex and wild-type antibody and antigen sequences and asked the models to redesign all three HCDRs while maintaining compatibility with the parental backbone and retaining or improving binding. We targeted approximately 50% sequence identity in each loop and held all other scFv residues fixed.
- Candidate selection. We used computational filtering and required substitutions at more than 40% of HCDR3 positions. (Section 8.3).

In the pure inverse folding task described in Section 3.2, the internal models readily recovered these antibodies' full wild-type sequences from their supplied backbones, consistent with our in silico results in Section 3.2.1.

The fraction of designs that passed the computational screen ranged from 0.14% for finotonlimab to 50.5% for tislelizumab (Figure 21). Despite this variation, median sequence identity to the parent across the three HCDRs combined was close to the requested 50% in all five families (51.7–52.0% among passing designs).

Wet-Lab Results

In the single-dose BLI screen, 179 of 272 designs from five parental antibody families screened positive for PD-1 or PD-L1 binding (65.8%). For PD-L1, 83 of 92 avelumab designs (90.2%) bound, and 48 of 92 cosibelimab designs (52.2%) bound. For PD-1, the hit rates were 90.3% for tislelizumab (28/31), 31.5% for serplulimab (17/54), and 100% for finotonlimab (3/3). These results show that the model can redesign antibody sequences with high hit rates. They also provide evidence against simple memorization as the sole explanation for the model's antibody inverse folding capabilities, as sequence redesign requires reasoning beyond memorization. Randomly substituting 50% of the residues in HCDRs is unlikely to preserve binding.

Finally, note tepresentative BLI screening traces are provided in Appendix D.2 (Figure A4A–E). Position-level HCDR sequence variation for tislelizumab, serplulimab, finotonlimab, avelumab, and cosibelimab is provided in Appendix D.3 (Figure A5A–E).

4 Capability Analysis: Structure Prediction

Protein structure prediction is a fundamental task in biology and drug discovery. The GPT-6 model family still struggles with end-to-end structure prediction, including monomer folding, although performance has improved drastically across model generations. On several related tasks across protein–protein docking, protein–ligand docking, side-chain packing, and local structure prediction, our models achieve performance comparable to that of state-of-the-art specialized models.

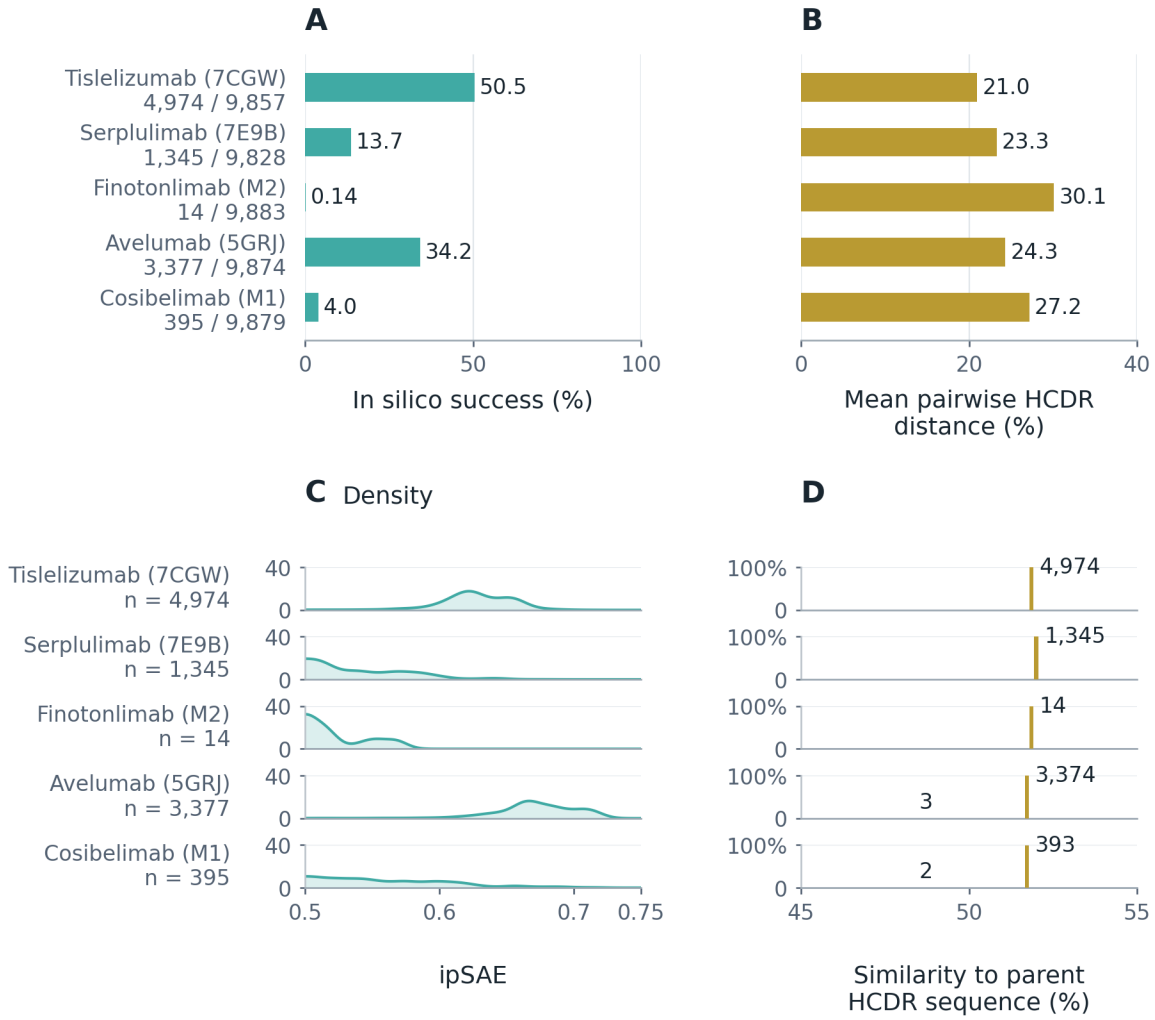


Figure 21. Computational selection of redesigned therapeutic antibodies. (A) Passing rates (passing/evaluable counts shown). (B) Within-context mean pairwise heavy-chain CDR (HCDR) sequence distance. (C) Density maps of ipSAE scores, (D) Parent HCDR sequence-identity frequencies, with counts labeled. C–D include all 10,105 passing records, including repeated sequences. Passing requires ipSAE $\geq$ 0.5, target-aligned binder RMSD $<$ 4 Å and each HCDR RMSD $<$ 2 Å. M1/M2 denote starting backbones.

4.1 Monomer Folding

Dataset and Settings. The task is to predict the coordinates of protein backbone atoms (N, Cα, C, and O) from an amino acid sequence. These backbone coordinates define the overall protein fold. We evaluated GPT in the standard setting and used ESMFold2 (Candido et al. 2026) for comparison.

Following Section 3.1, we use subsets of two datasets: experimentally determined structures from the RCSB Protein Data Bank (RCSB PDB; Berman et al. 2000) and predicted structures from the AlphaFold Protein Structure Database (AFDB; Varadi et al. 2022). For this evaluation, we exclude protein complexes and consider only monomeric protein structures.

Evaluations. We primarily use TM-score (Zhang and Skolnick 2004) and local Distance Difference Test (lDDT; Mariani et al. 2013) to evaluate the quality of predicted protein structures. Together, these metrics provide complementary assessments of global fold accuracy and local structural quality.

- TM-score (higher is better): Measures global structural similarity between the predicted and reference structures, with values ranging approximately from 0 to 1. A TM-score above 0.5 generally indicates that two structures share a similar overall fold (Xu and Zhang 2010), while scores above 0.8 indicate very high structural similarity.
- lDDT (higher is better): Measures the accuracy of local interatomic distances without requiring structural superposition. It captures how well the predicted structure preserves local geometric relationships relative to the reference structure.

4.1.1 Results

4.1.2 Internal Models Can Fold Nontrivial Protein Structures

At the highest evaluated budget, the internal model's mean lDDT was 0.40 on RCSB PDB and 0.38 on AFDB. Mean TM-scores were 0.38 and 0.25, respectively. Both metrics remained well below the ESMFold2 references (Figure 22A–D).

In a separate set of scored predictions, some structures had both a backbone RMSD below 5.0 Å and a TM-score of at least 0.6. Figure 22E–G shows the score distributions, and Figure 22H illustrates four selected cases. Such successful predictions were rarely observed with earlier models.

4.1.3 CDR Structure Prediction Performance Is High

Our internal models perform substantially better on local structure prediction tasks when part of the protein structure is provided. We evaluated structure prediction of each flexible loop of the antibody complementarity-determining regions (CDRs), given the antibody sequence and the surrounding framework structure.

At the highest evaluated budget, median backbone RMSD for the internal model ranged from 0.71 to 1.58 Å across five of the six CDRs and reached 2.46 Å for HCDR3 (Figure 23B). ESMFold2 had lower median RMSD on five of the six loops.

4.2 Multimer Rigid-Body Docking (Including Antibody–Antigen Docking)

Dataset and Settings. We ask our models to predict the relative position and orientation of two proteins to reconstruct their native complex. We evaluated two variants of this task. In the first, models received all-atom structures, including both backbone and side-chain coordinates. In the more challenging second variant, models received only backbone coordinates.

We used multimeric protein structures from the RCSB Protein Data Bank (Berman et al. 2000) and an antibody-focused dataset from SAbDab (Dunbar et al. 2014). For each complex, we separated the interacting proteins and asked the models to recover their relative positions and orientations.

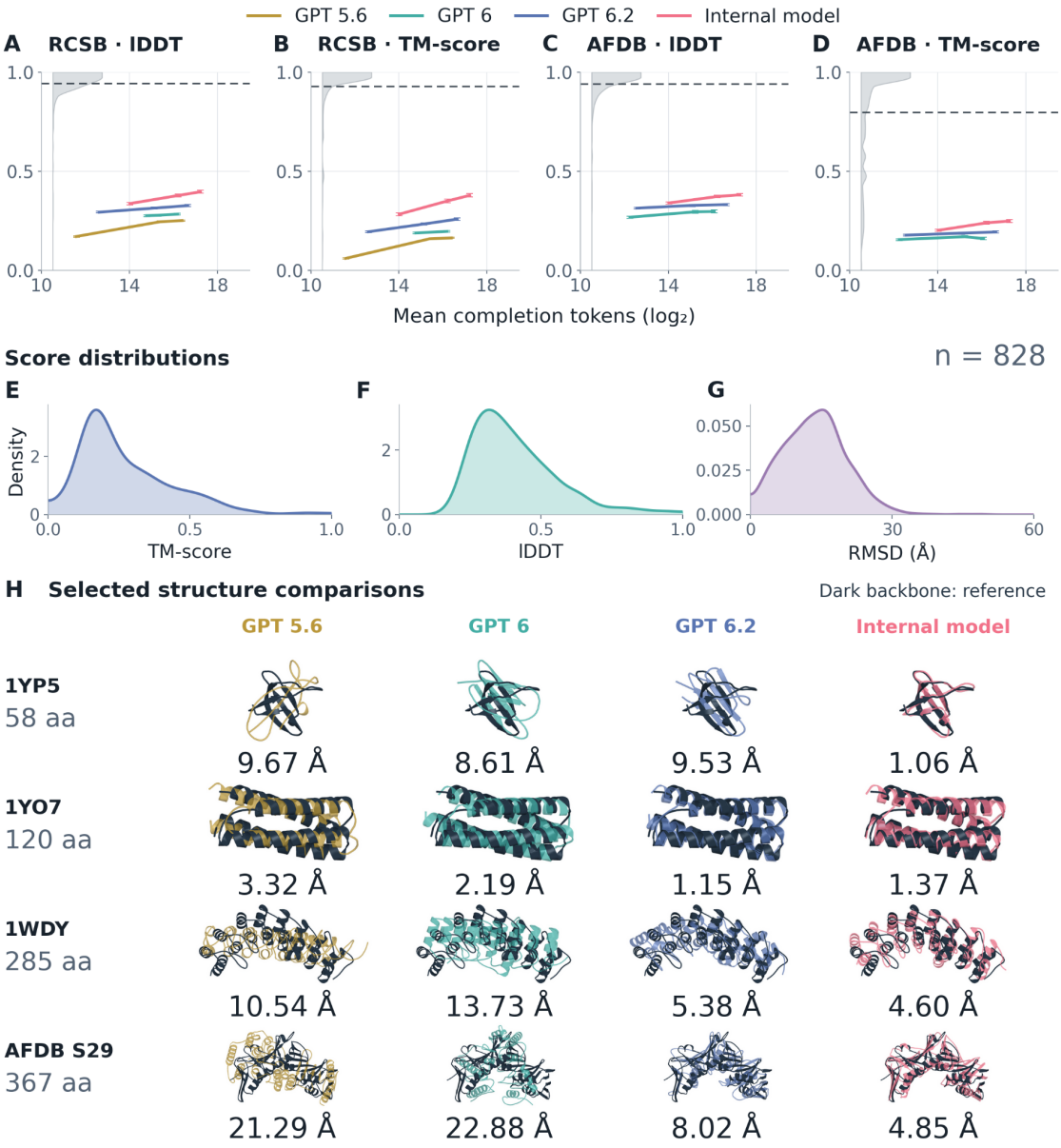


Figure 22. Monomer structure prediction across computational budgets and selected structures. (A–D) Backbone lDDT and globally Cα-aligned TM-score for RCSB and AFDB tasks. Curves connect mean scores across model-specific budget settings; error bars indicate one standard error across completions. The x-axis is $\log_2$ of mean completion-token counts. Gray distributions show ESMFold2 predictions; dashed lines mark their mean scores: 0.943 for RCSB lDDT (A), 0.928 for RCSB TM-score (B), 0.941 for AFDB lDDT (C), and 0.796 for AFDB TM-score (D). (E–G) Normalized kernel-density estimates of TM-score, backbone lDDT, and backbone RMSD from a separate frozen snapshot of scored root samples. (H) Four illustrative cases span 58–367 residues and experimental RCSB or predicted AFDB references. Colored predictions overlay dark reference structures; labels give RMSD (Å) after global N/CA/C/O backbone alignment. Examples were selected from the existing gallery, are not a population sample, and use model-specific budgets. The companion gallery retains all 22 cases. For more samples, refer to Appendix Figure E1.

Because the individual structures were extracted from experimentally determined complexes, they retain their bound (holo) conformations. The benchmark therefore approximates bound-state rigid-

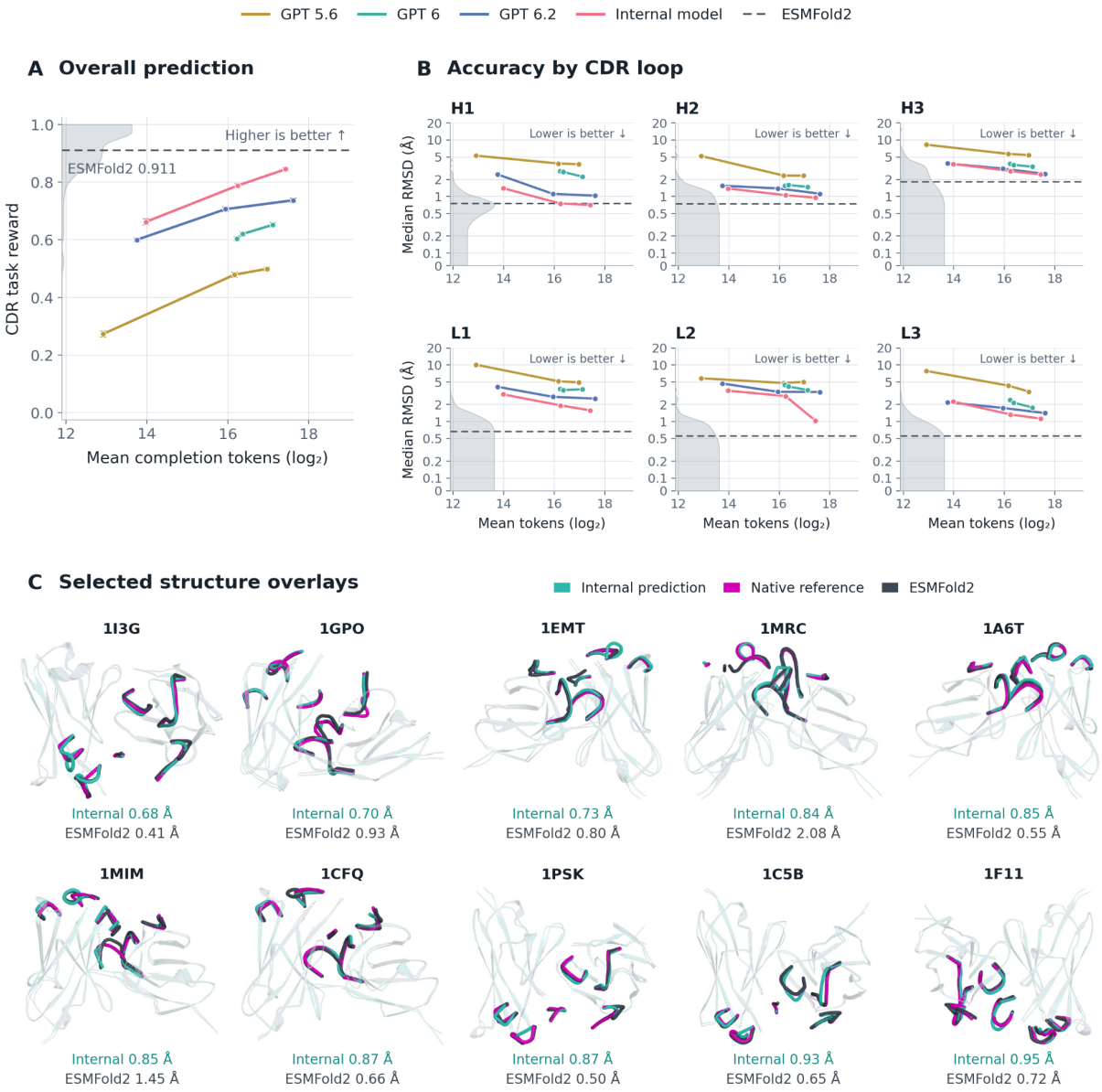


Figure 23. Antibody CDR structure prediction from sequence and framework coordinates: (A) overall task reward and (B) median backbone RMSD for each heavy- and light-chain CDR loop versus $\log_2$(mean completion tokens). Dashed lines and gray distributions show ESMFold2 references. (C) Selected overlays compare internal-model predictions (teal), native loops (magenta), and ESMFold2 (dark gray); labels give mean RMSD across all six CDRs after framework alignment.

body docking and does not capture conformational changes between the unbound (apo) and bound states.

Evaluations. We use DockQ (Basu and Wallner 2016) as the primary evaluation metric. DockQ is a continuous metric ranging from 0 to 1 that measures the similarity between a predicted protein–protein complex and its experimentally determined reference structure. Higher DockQ scores indicate more accurate docking configurations. Following the standard DockQ classification, scores

can be interpreted approximately as follows: (a) DockQ ≥ 0.23: acceptable docking quality, (b) DockQ ≥ 0.49: medium-quality docking, and (c) DockQ ≥ 0.80: high-quality docking, indicating near-native complex structures.

4.2.1 Results

At the highest evaluated budget, the internal model achieved mean scores of 0.96 DockQ for antibody–antigen docking and 0.95 GlobalDockQ for multimer docking when side chains were supplied. With backbone coordinates alone, the corresponding means were 0.81 and 0.85 (Figure 24A–D). Both settings used chains in their bound conformations, so these scores measure how well the model reassembled a complex from known structures. This distinction also matters for specialized structure prediction models, where strong monomer folding performance does not necessarily translate into accurate multimer prediction because modeling protein–protein interactions poses additional challenges (Yin et al. 2022).

4.3 Blind Docking With Small Molecules

Dataset and Settings. The task is blind protein–ligand docking, where models must predict the binding pose of a small molecule given the protein structure and ligand SMILES only. A 3D structure is not provided for the ligand. For the protein, we consider two variants: (1) all-atom structures, including both backbone and side-chain coordinates, and (2) backbone-only structures, where side-chain coordinates are removed. The latter represents a more challenging setting, as models must infer the missing side-chain geometry while predicting the ligand binding pose. We compare GPT with DiffDock (Corso et al. 2023, 2024).

We evaluate performance on two established benchmarks: PoseBusters (Buttenschoen et al. 2024) and SCARPiNo (Scarpino et al. 2018). PoseBusters evaluates noncovalent protein–ligand docking, whereas SCARPiNo focuses on covalent docking, where ligands form covalent bonds with specific protein residues. Models must perform blind docking, a setting in which no binding pocket information or initial ligand coordinates are provided. For SCARPiNo, we additionally provide an explicit cysteine–ligand attachment pair, specifying the protein residue and ligand atom involved in covalent bond formation.

Evaluations. We use ligand RMSD as our primary evaluation metric. It measures the geometric difference between the predicted and experimentally determined ligand poses after aligning the corresponding receptor structures. Lower RMSD values indicate more accurate docking predictions.

4.3.1 Results

With receptor side chains supplied, GPT-6, GPT-6.2 and the internal model each outperformed DiffDock-L on both PoseBusters and Scarpino at their highest evaluated budgets. All three had lower median ligand RMSDs and a higher fraction of poses within 2 Å of the reference (Figure 25A–B). The internal model had the lowest median RMSDs, 0.34 Å on PoseBusters and 0.43 Å on Scarpino, compared with 1.89 Å and 1.27 Å for DiffDock-L. Even without side-chain coordinates, the internal model predicted ligand poses from SMILES and the receptor backbone with median errors below 1 Å on both benchmarks (0.67 Å and 0.56 Å, respectively).

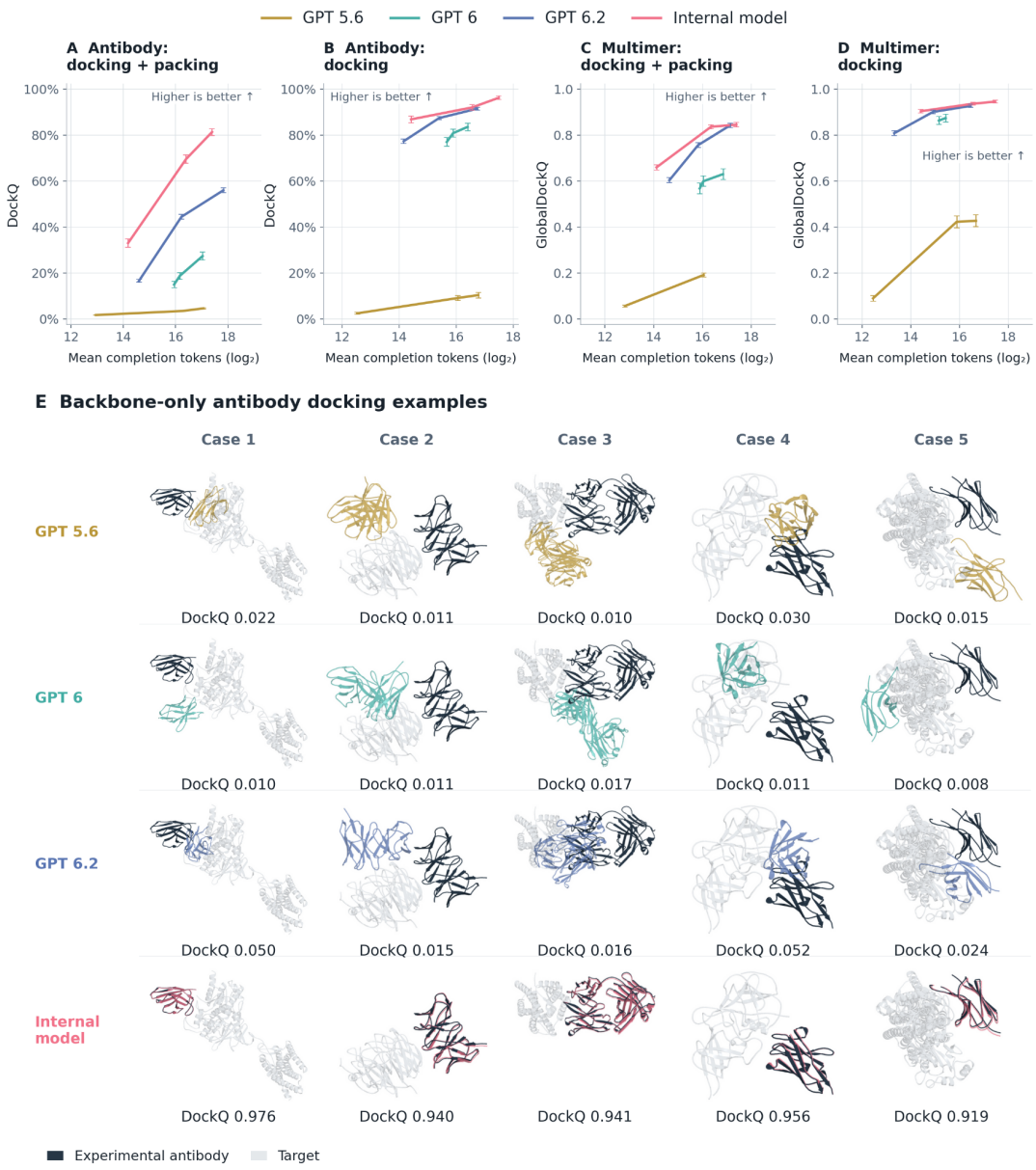


Figure 24. Protein–protein docking performance and representative antibody docking structures. (A–B) Mean DockQ for antibody–antigen docking with and without side-chain packing, respectively. (C–D) Mean GlobalDockQ (Mirabello and Wallner 2024) for multimer docking with and without side-chain packing, respectively. Lines connect mean scores at the evaluated inference budgets against $\log_2$ mean completion tokens; error bars show standard errors. Higher scores indicate better agreement with experimental complexes. (E) Five source-selected backbone-only antibody docking cases compared across all four models, with exact DockQ values shown to three decimals. Predicted antibodies use the model colors; experimental antibodies are dark blue-gray and targets are light gray. All four models within each case share the same camera, crop, and scale.

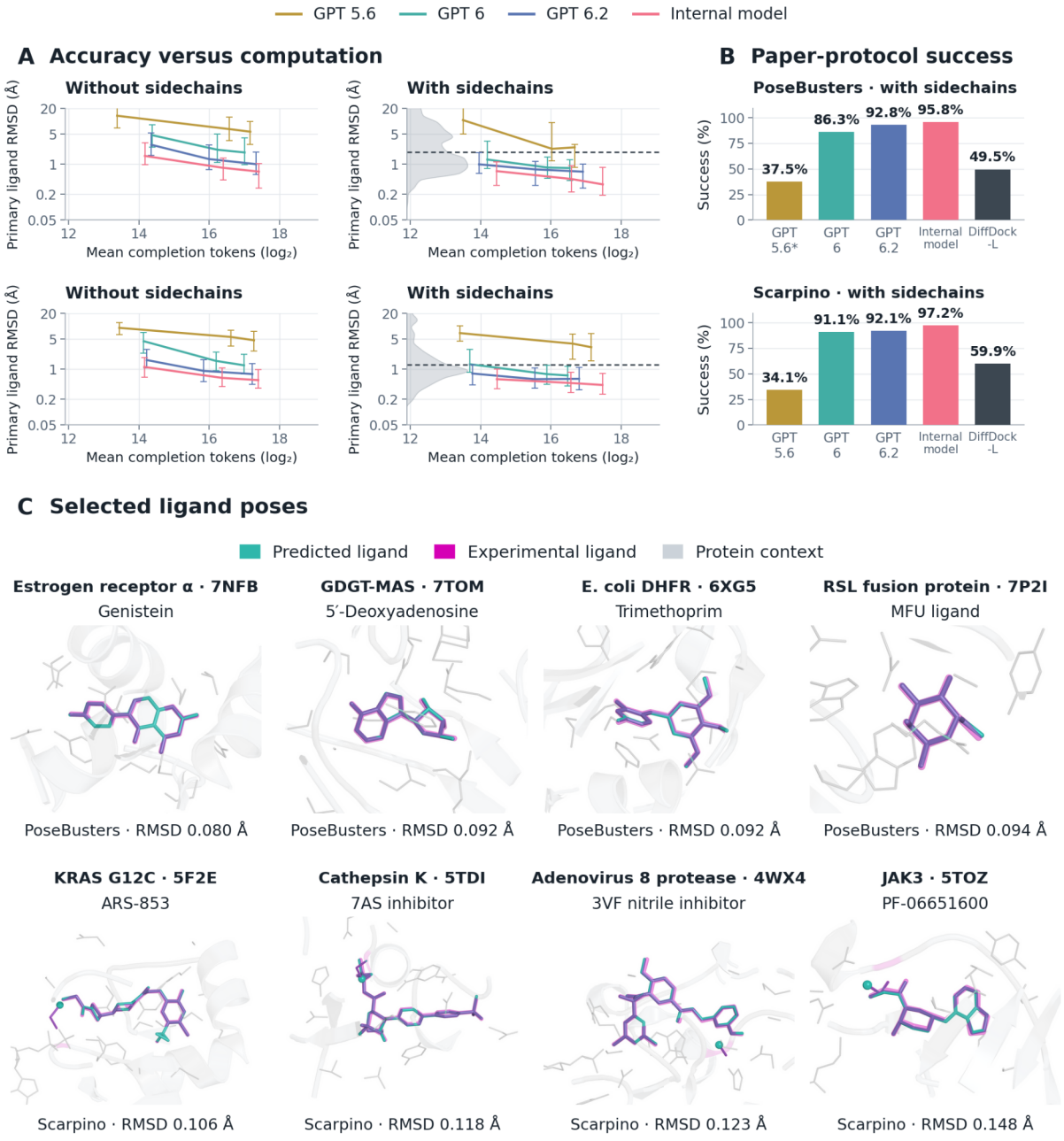


Figure 25. Blind docking with small molecules. (A) Primary ligand RMSD versus $\log_2$ mean completion-token count on PoseBusters (top row) and Scarpino (bottom row), without and with side chains; lines connect medians and error bars show interquartile ranges. Lower RMSD is better. Gray density profiles and dashed lines show the original DiffDock-L distributions and median RMSDs: 1.89 Å for PoseBusters with side chains (top middle) and 1.27 Å for Scarpino with side chains (bottom middle). (B) Paper-protocol success on PoseBusters and Scarpino with side chains at each model's highest selected budget; success is paper-protocol ligand RMSD $\leq$ 2 Å. (C) Eight selected internal-model poses from PoseBusters (top row) and Scarpino (bottom row), with predicted ligands in teal, experimental ligands in magenta, and protein context in gray.

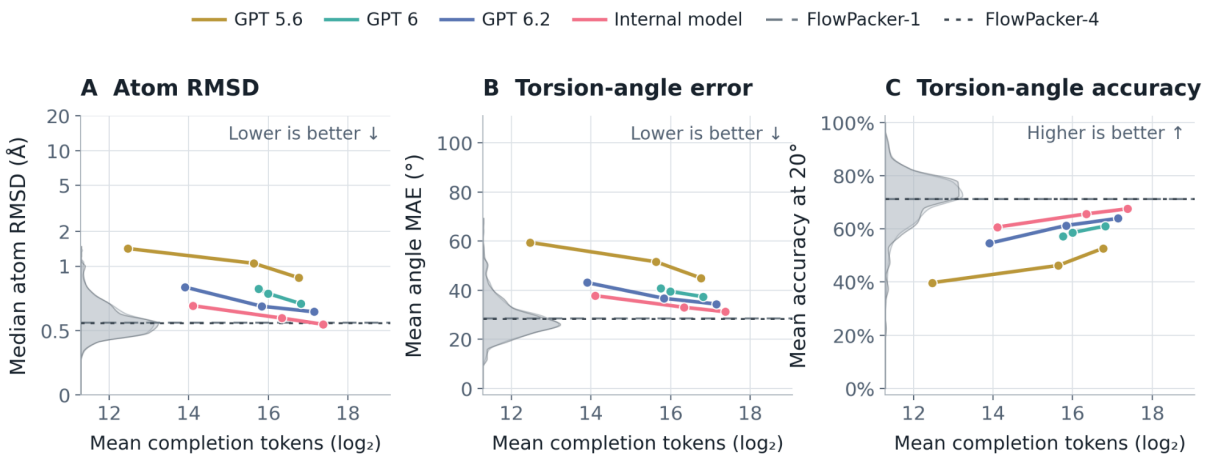


Figure 26. Side-chain packing accuracy versus $\log_2$(mean completion tokens). Panels show (A) median side-chain heavy-atom RMSD, (B) mean absolute torsion-angle error, and (C) mean torsion-angle accuracy within 20°; lower errors and higher accuracy indicate better performance. Dashed lines and gray profiles show FlowPacker-1 and FlowPacker-4 references and score distributions; the RMSD axis is linear below 1 Å and logarithmic above.

4.4 Side-Chain Packing

Dataset and Settings. In the side-chain packing task, models receive protein backbone structures and amino acid sequences and predict the corresponding side-chain atomic coordinates. We evaluate the models on several protein structures from the RCSB PDB (Berman et al. 2000). We compare their performance with that of FlowPacker (Lee and Kim 2025), a representative specialized deep learning model for side-chain packing.

Evaluations. We evaluate side-chain packing performance using the following metrics (Lee and Kim 2025):

- Side-chain torsion angle MAE (lower is better): The mean absolute angular error for the four side-chain dihedral angles ($\chi1$, $\chi2$, $\chi3$, and $\chi4$), measured in degrees. We calculate the minimum periodic angular difference between predicted and reference torsion angles. Lower values indicate better performance.
- Side-chain torsion angle accuracy (higher is better): The fraction of predicted side-chain dihedral angles ($\chi1$, $\chi2$, $\chi3$, and $\chi4$) that fall within a specified angular threshold of the reference values (commonly 20°). Higher values indicate better performance.
- Side-chain atom RMSD (lower is better): The root-mean-square deviation between predicted and reference side-chain heavy-atom coordinates, calculated with the backbone fixed and accounting for chemically equivalent atoms where appropriate. Lower values indicate more accurate side-chain conformations.

4.4.1 Results

At its highest evaluated budget, the internal model approached FlowPacker's accuracy in side-chain packing, with the remaining gap concentrated in torsion angles (Figure 26). Median side-chain heavy-atom RMSD was nearly identical for the internal model and FlowPacker-4 (0.55 and 0.56 Å, respectively). The internal model had a higher mean absolute torsion-angle error (31.2° versus 28.4°) and lower accuracy within 20° (67.5% versus 71.2%). These results suggest that the model captures many of the geometric constraints on side-chain placement.

5 Capability Analysis: Variant Effect Prediction

Variant effect prediction estimates how amino acid substitutions affect protein properties, such as stability, binding affinity, and biological function (Meier et al. 2021; Notin et al. 2023). Surprisingly, GPT-6.2 and our internal models demonstrate strong capabilities in variant effect prediction, outperforming a commonly used zero-shot baseline based on ESM2.

Dataset and Settings. We use four evaluation datasets constructed from deep mutational scanning (DMS) and related high-throughput experimental measurements of protein sequence variants:

- ProteinGym (Notin et al. 2023): We use the DMS substitution component of ProteinGym, which groups assays into five broad categories: activity, binding, expression, organismal fitness, and stability.
- Mega-Scale (Tsuboyama et al. 2023): A dataset of thermodynamic folding stability across variants of small natural and de novo designed protein domains.
- BindingGYM (Lu et al. 2024): A curated collection of DMS assays measuring the effects of sequence variants on protein–protein interactions, including antibody–antigen, receptor–ligand, signaling protein, and protein–peptide interactions. Systems in our evaluation include 4D5–HER2, ACE2–SARS-CoV-2 RBD, KRAS with signaling partners, and BH3 peptides with Bcl-xL or Mcl-1.
- AlphaBio (A-Alpha Bio 2025; Agarwal et al. 2025; Younger et al. 2017): A collection of experiments estimating protein–protein binding affinity using AlphaSeq, a yeast mating assay. Our evaluation uses variants of VHH72, PP489, pembrolizumab-scFv, and trastuzumab-scFv tested against fixed protein partners, including coronavirus receptor-binding domains, TIGIT, PD-1, and HER2.

Note that these datasets are not necessarily mutually exclusive. For example, ProteinGym includes data from the Tsuboyama Mega-Scale dataset (Notin et al. 2023).

We formulate variant effect prediction as a ranking task, following Arora et al. (2026). Models rank 50 protein variants by their predicted values for the experimentally measured property. Unlike the evaluations in other sections, we do not allow GPT models to use coding tools in this benchmark. For each task, we provide the wild-type sequence, the 50 variant sequences, and basic context, including the assay type and the property being measured. For simplicity, we restrict the evaluation to substitution variants, excluding insertions and deletions. As a baseline, we compare against ESM2-650M (Lin et al. 2023; Notin et al. 2023), a widely used protein language model for zero-

shot variant effect prediction. Specifically, we use ESM2's sequence likelihood-based scores to rank variants, following standard zero-shot variant effect prediction approaches (Meier et al. 2021).

Evaluations. We use Spearman's rank correlation coefficient between the predicted rankings and the ground-truth experimental rankings as our primary evaluation metric.

5.1 Results

GPT-6.2 and our internal models achieve substantially higher Spearman correlations than the zero-shot ESM2 baseline across the evaluated datasets (Figure 27). This trend holds across several assay categories in ProteinGym and for most protein targets in BindingGYM.

Our models may have encountered related protein sequences or experimental measurements during training, so their predictions may not meet the same zero-shot criteria as the ESM2 baseline. Nevertheless, the consistently strong performance across datasets and assay types suggests that GPT models have learned patterns relevant to protein sequence–function relationships.

6 How Does GPT Address Problems in Protein Biology?

Across the examples in Appendix E, the models often proposed a candidate, checked it with calculations they implemented in code, and revised it when the checks exposed a problem. The examples below show what guided those decisions in each task.

6.1 Structure Design

Across minibinder design, motif scaffolding, and topology-conditioned backbone design, the models translated structural requirements into explicit geometric plans and used custom code to construct and refine candidate backbones. Geometric and chemical checks guided candidate selection and local repairs, with temporary models of side chains or cofactors sometimes used to assess packing and proposed interactions. This iterative process allowed the models to revise both their candidate structures and the assumptions guiding their construction.

6.1.1 Minibinder Design

Models commonly approached binder design through three stages: identifying a suitable arrangement at the target surface, constructing and comparing candidate scaffolds, and refining the selected structure. See Appendix E.1.1 for details. They began by inspecting the specified target patch and defining a local coordinate frame for binder placement. In one IL-7Rα trajectory, surface inspection supported the model's hypothesis that a recognition helix could pack along a hydrophobic ridge near Tyr159, motivating tests of helix caps that controlled how the backbone turned away from this interface.

Once an arrangement had been proposed, the models built helices and β-hairpins from peptide bond lengths, bond angles, and backbone torsions, then joined these elements with short connectors. For some PD-L1 examples, the model searched a loop library for connectors that matched both the positions and orientations of adjacent helix endpoints. Local tests helped refine these construction

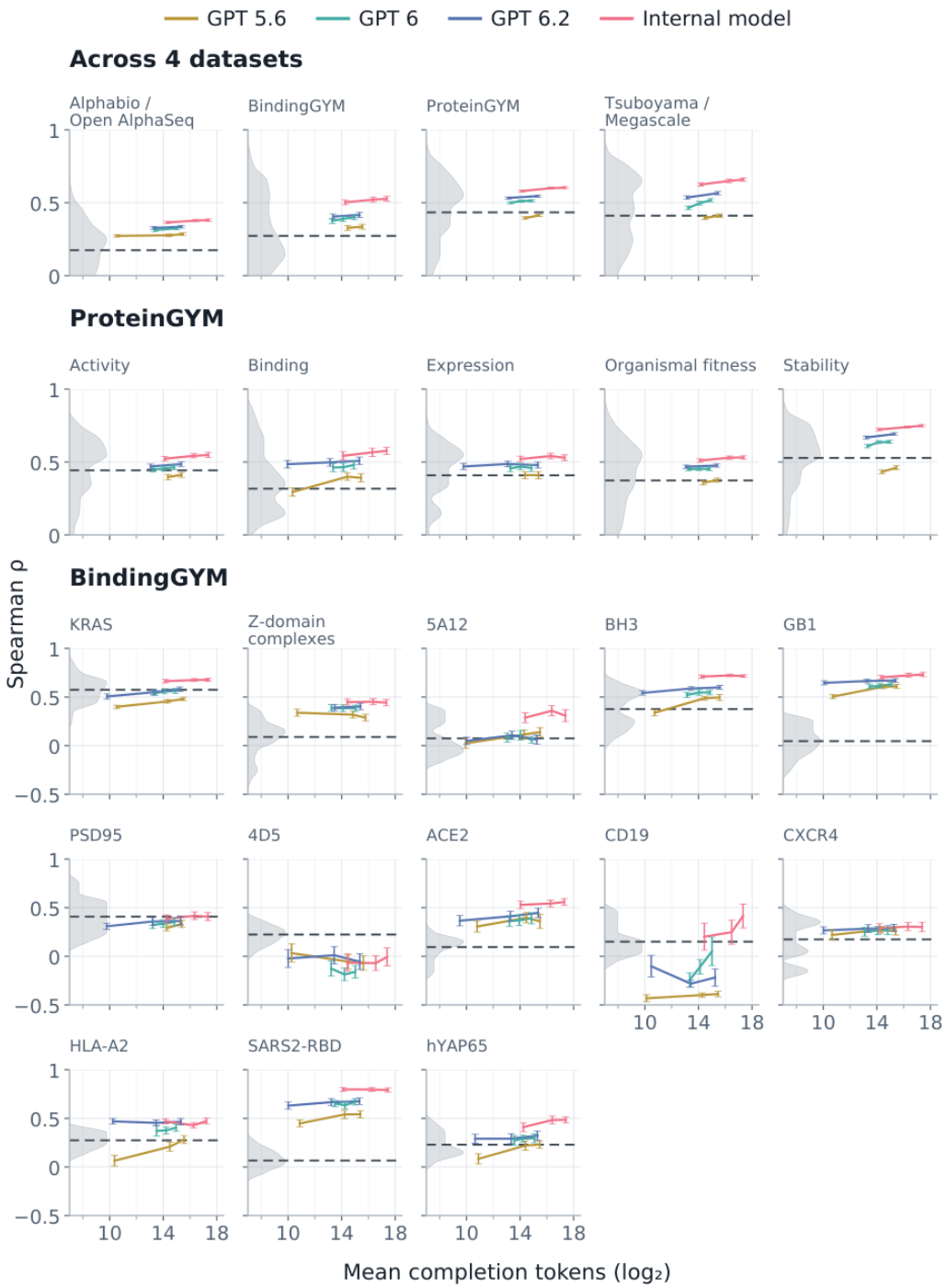


Figure 27. Variant-effect prediction measured by Spearman correlation between predicted and experimental variant rankings versus $\log_2$(mean completion tokens). Results are shown across four datasets, by ProteinGym assay category, and by BindingGYM target category. Points show means ± 1 standard error; dashed lines and gray distributions show ESM2 means and score distributions, with higher correlations indicating better ranking performance.

rules: in some VEGF-A examples, an initial turn test accepted no fragments, but a repeated test returned accepted alternatives after the model revised both the acceptor offset and backbone torsions. The resulting scaffolds were compared using custom criteria for target contacts, steric overlap, internal packing, torsional strain, and polar interactions.

The final stage combined sequence choices with structural refinement. In several trajectories, the models alternated between amino acid and side-chain conformation choices and continuous coordinate adjustments (even though they were not explicitly asked to construct side chains), allowing sequence and structure to accommodate one another. In some VEGF-A examples, the model rejected a Gln62 option despite a 2.90 Å target contact because it judged the side-chain angle strained, and returned to a Lys62 candidate. For TrkA, it selected candidate 34 over a lower-scoring alternative based on its stated helix/core preference, then reduced a custom steric-overlap diagnostic from 0.337 to 0.101 Å through local refinement. Before submission, the models checked peptide continuity, preservation of the supplied target, and interface contacts, completing a process in which structural hypotheses were repeatedly tested and revised through geometric and chemical calculations.

6.1.2 Motif Scaffolding

The model builds a complete protein backbone by choosing a supporting fold, connecting the supplied fragments, and refining the surrounding geometry and packing. See Appendix E.1.2 for the details. The motif coordinates act as fixed anchors; some residue identities are also fixed, while other positions remain designable. The model examines sequence patterns, backbone angles, and side-chain positions to decide whether helices, sheets, or a mixture can support the motif. For example, the four helical fragments in 1BCF (Frolow et al. 1994) guide construction of a compact helical bundle.

The model then searches for an order and length for the connecting segments that will join every fragment within the required total chain length. In 1BCF, this search selects the fragment order D–C–A–B and connects turns of 3, 5, and 3 residues, producing a 125-residue chain. Custom numerical code places backbone atoms using bond lengths, angles, and rotations, then refines candidate chains by penalizing atomic overlaps and distorted peptide geometry. Local checks prompt further repairs, such as bringing two computed copies of a junction into agreement or correcting an atom index in a geometric calculation.

Temporary side-chain models test whether the scaffold leaves room for packing and can accommodate proposed polar or sulfur contacts (again even though they were not explicitly asked to construct side chains). In 5WN9 (Fedechkin et al. 2018), the model fits a new cysteine side chain toward a fixed cysteine using an S–S distance target of 2.038 Å, then includes the fitted atoms in the saved structure. Before submission, it verifies residue count, chain continuity, required atoms, and exact preservation of the supplied motif coordinates. Later sequence design and refolding assess backbone and motif recovery; biological function remains untested.

6.1.3 Topology-Conditioned Backbone Design

In general, the model turned a structural classification and a requested chain length into a protein backbone through a sequence of planning, coordinate construction, numerical search, and geometric checks. See Appendix E.1.3 for details. It began by interpreting the CATH label as a possible

arrangement of helices and β-strands, then decided their order along the chain and how they should pack together. For example, the YaeB-like design used a five-stranded β-sheet with two helices, while the ubiquitin-like design used a β-sheet wrapped around a helix. These fold interpretations provided working hypotheses that narrowed the structures the model needed to explore.

The model then wrote code to turn its plan into three-dimensional coordinates using bond lengths, bond angles, and backbone torsion angles, which control how the chain bends and twists. It constructed helices and strands, positioned them relative to one another, and generated or searched for loops that could connect them. The planned arrangement also became a set of numerical constraints: in the ubiquitin-like example, the model specified which strands should pair and which backbone atoms should form hydrogen bonds. In the Gal4-like example, it introduced temporary zinc and sulfur positions to guide the geometry, then omitted these temporary atoms from the submitted backbone. It searched alternative conformations using scores that balanced the intended contacts against atomic overlap, unfavorable local geometry, and packing.

As a final step, the model inspected the candidates using concrete measurements, including interatomic distances, torsion angles, and contact counts, and used these checks to identify local problems. In the Gal4-like example, short sulfur contacts raised a steric concern, leading the model to change a helix boundary and a cysteine orientation before checking the metal-site geometry again. The final choice could combine numerical ranking with inspection of the resulting structure, including tradeoffs between competing geometric objectives. Finally, the model wrote the selected backbone to a PDB file and checked properties such as residue count, required backbone atoms, chain continuity, and file formatting.

6.2 Sequence Design

Across protein and antibody inverse folding, the models combined prior sequence knowledge with calculations tailored to the supplied backbone. Candidate residues and motifs were evaluated through checks of local geometry, side-chain packing, and chemical interactions, which could prompt revisions to initial choices. The final sequences reflected a balance between sequence-based expectations and structural scores, including cases where the two favored different residues.

6.2.1 Inverse Folding

The model generates amino-acid sequences from supplied protein backbones by combining protein-family hypotheses, candidate sequence fragments, and structural calculations. For details, see Appendix E.2.1.

The process begins by examining features such as helix and strand arrangement, chain similarity, and local backbone geometry. These observations can suggest a protein family or fold, which the model uses to enter a candidate sequence, select a sequence prefix, or propose characteristic residues at particular sites. The proposal is then made compatible with the input: candidate sequences are cropped or aligned to the backbone length, separate blocks are joined, and selected residue identities are fixed or given higher preference. For example, some data points combine a stored sequence prefix with a separately constructed tail, while others impose candidate cysteine pairs or metal-coordinating residues as starting assumptions.

The model then evaluates or refines these choices using calculations of side-chain packing, hydrogen

bonding, solvation, residue contacts, or local geometric fit. In several examples, it averages residue preferences from multiple calculations and selects the highest-scoring amino acid at each position; in others, most letters come from an entered template and the calculations assess only selected sites. Final sequences are assembled through explicit operations such as joining blocks, removing terminal residues, copying a shared sequence across chains, or replacing individual letters. These checks support the construction of a candidate, while protein-family assignments and imposed motifs remain hypotheses; the final evaluation separately measures exact agreement with the native sequence.

6.2.2 Inverse Folding With Antibodies

Examples of how the model chose residues for masked antibody loops appear in Appendix E.2.2. It first used the visible antibody framework and familiar sequence motifs to suggest initial candidates, then examined the backbone geometry to narrow the choices at individual positions. For example, in 1DL7 (Brown et al. 2000), adding a trial side-chain carbon at heavy chain position 53 produced a clash with the backbone, supporting glycine, which lacks that carbon. Geometric checks and motif preferences guided further substitutions, changing the second heavy chain loop from WAGGS to WGDGS and finally WGDGN.

To compare candidates, the model wrote custom code that sampled side-chain orientations and scored their packing and contacts with nearby atoms. In 3S35 (Franklin et al. 2011), it fitted a scoring function using visible residues, checked its predictions at positions excluded from fitting, and repaired a memory error in the code linking Python and C before rerunning the search. Some calculations also accounted for interactions between candidate choices: in 4ZD3 (Chen et al. 2015), the model repeatedly updated candidate weights based on compatibility between neighboring positions, so a choice at one position could change the ranking at another. The structural environment supplied further evidence; in 5E08 (Shao et al. 2016), for example, the model evaluated contacts between proposed side chains and water molecules or RNA phosphates already present in the structure.

These calculations could prompt the model to reconsider an earlier assumption, while sequence motifs continued to influence the final choice. In 5T1D (Sathiyamoorthy et al. 2016), a clash led the model to remove a restriction that allowed only proline at one position, reopening the search to other amino acids. It ultimately selected methionine based on a motif preference, despite tyrosine having the better local score. This example illustrates how the final sequence could reflect both structural calculations and sequence-based expectations, including cases where the two favored different residues.

6.3 Structure Prediction

Across structure prediction and docking, the models used sequence patterns, biological context, and geometric constraints to narrow the search for plausible structures. They then generated and compared candidates with custom code, using local chemistry and agreement among candidates to guide refinement. The available constraints shaped this process: fixed antibody frameworks and supplied docking partners restricted the search, whereas monomer prediction required the models to propose a global fold from sequence.

6.3.1 Monomer Structure Prediction

The model used sequence clues to infer the overall fold in the examples in Appendix E.3.1. It first used repeated sequences, recognizable motifs, and recalled features of protein families to decide how helices and sheets might be arranged. In 6I3B (Mylemans et al. 2020), six identical repeats of 42 residues suggested a sixfold structure, so the model built one unit and copied it using 60° rotations. In 1WDY (Tanaka et al. 2004), an ankyrin repeat hypothesis organized the chain into eight units, each containing a pair of helices. For the elongated 3TIP structure (Gruszka et al. 2012), the model built overlapping fragments, rotated and translated them until their shared regions matched, and used guide points along the long axis to keep the assembled chain extended.

The model then expressed its structural plan as numerical rules that custom code could evaluate, including preferred backbone angles, strand pairings, hydrogen-bond distances, and penalties for overlapping atoms. Several examples favored backbone N–O distances near 3 Å to support hydrogen bonds between strands. In 2WJH (Köster et al. 2009), the model used assumed contacts with a nucleotide and magnesium to bring distant parts of the sequence into one pocket: it targeted 2.98 Å between Lys16 and a phosphate oxygen, 2.12 Å between Ser17 and magnesium, and 2.89 and 2.92 Å for two Asp119–guanine contacts. Its search code generated alternatives by rotating backbone bonds, shifting which residues paired across strands, and recombining previously built fragments; the 2WJH example ran 12 searches of 562 trials before ranking candidates.

Finally, the model used agreement among candidates to guide refinement, applying stronger restraints to consistent core regions and weaker restraints to variable loops. It also inspected local geometry when choosing among candidates: in 1G2B (Berisio et al. 2001), it selected a less strained loop around proline 20 despite that candidate having a less favorable internal score. In 4Y0A (Sutton et al. 2015), it added a preference for an Arg143 contact that overrode the earlier score advantage of an Arg137 alternative. Before submission, the model checked that the chain remained connected, identified unusually close atomic contacts, adjusted local geometry, and wrote the complete N, Cα, C, and O backbone for comparison with the withheld reference structure.

6.3.2 CDR Structure Prediction

Examples of CDR structure prediction appear in Appendix E.3.2. The fixed framework supplied the central constraint: each missing loop had to connect two existing chain ends. Using the framework PDB and Chothia numbering (Chothia and Lesk 1987), the model located each gap and used its flanking N, Cα, and C atoms to define the required endpoint positions and chain directions. Within these constraints, it sampled backbone angles (ϕ and ψ) according to residue type; in 1GPO (Schiweck and Skerra 1997), it also specified a cis peptide preference before L3 Pro95 ($\omega = 0°$).

Candidate construction was guided by both geometric compatibility and chemical scoring. Some searches grew shorter fragments from both ends, matched their positions and orientations at an internal junction, and adjusted backbone rotations to reduce closure error. In 1CIC (Bentley et al. 1990), the H3 generator was configured for 120,000 draws and 7,200 entries for refinement, while the execution log directly reported 3,798 retained candidates. For detailed scoring, the model temporarily constructed side chains and assessed atomic crowding, packing, hydrogen-bond geometry, charge interactions, and geometric estimates of water access. In 1MIM (Mikol 1996), for example, 86 probe directions at radii of 2.95 and 3.24 Å from polar atoms were used to estimate

access to water.

Candidate selection and refinement combined numerical scores with structural judgment and explicit checks of the resulting geometry. Random score perturbations encouraged exploration, and diversity filters retained different loop shapes; clustering or weighted coordinate averages then summarized agreement among candidates. In 2C1O (Parkkinen et al. 2006), the model favored a proposed Asn–Gly turn with Gly30D at $\phi \approx +90^\circ$, even though that candidate had a worse internal score. Selected loops were refined individually or jointly, with additional distance penalties targeting close contacts and poor connections to the framework. In 1PSK (Pichla et al. 1997), reoptimization after several score changes, including backbone/Cβ repulsion, increased the Asn52 N–Gly56 O separation from 2.62 to 2.87 Å. In examples with explicit final validation, the model checked the written backbone PDB for residue identities, framework preservation, C–N distances between consecutive residues, and unusually short nonadjacent contacts.

6.3.3 Protein Docking

The model assembled protein complexes from separate backbone structures by solving a sequence of smaller problems. See Appendix E.3.3 for the details. It first proposed which surfaces should meet, then rotated and translated each chain without changing its backbone, rebuilt missing side chains, and checked the resulting geometry.

Across these examples, biological hypotheses and geometric constraints narrowed the docking search to a manageable set of candidate poses. In the 1BGS example, a barnase–barstar hypothesis (Guillet et al. 1993) focused the search on complementary charged patches, with basic Arg/Lys residues facing acidic Asp/Glu residues. A coarse search used 980 starts, followed by ranking 68 detailed poses. In the 4Y0Y example (Ye et al. 2015), the model used a trypsin–inhibitor hypothesis and aimed Lys15 toward an acidic pocket and its carbonyl oxygen toward the oxyanion hole, a region that stabilizes negative charge. Symmetry simplified the geometry: the 2DWS trimer example (Jacobson et al. 2007) fixed a 120° rotation and optimized two translations, while the 6REX membrane example (Kovalev et al. 2019) used the ends of seven proposed helices to define a membrane frame before adjusting the axis and tilt of a fivefold assembly. Depending on the example, candidate placements were compared by matching protein shapes on grids or by repeated local optimization, with attractive contact terms and penalties for atomic overlap.

After selecting promising placements, the model sampled rotamers (i.e., common conformations of side chains) and relaxed local geometry using terms for atomic overlap, charge, solvation, and hydrogen bonding. Some examples also added configured disulfide targets or virtual metal and cofactor sites to represent hypothesized chemistry. Chemical inspection could override a small score advantage: in an intermediate comparison from 6REX, the model favored a Ser70 orientation (PDB Ser68) with a slightly worse custom score because it reduced oxygen crowding. The 2WYT example, guided by a SOD1 hypothesis (Antonyuk et al. 2010), compared nearly tied Ile113 packing states, while the 4FP5 example, guided by a toxin pentamer hypothesis (Cody et al. 2012), found that different objectives favored different copy orders.

6.3.4 Blind Docking

In blind docking, clues about the binding site and ligand chemistry helped the model decide which poses to test. For details, see Appendix E.3.4.

The initial clues differed across examples: the Ala80 kinase hinge guided placement in 6ZCY (Read 2020), magnesium ions and the ANP cofactor provided geometric targets in 7UY4 (Reeve et al. 2022), and supplied cysteine attachment sites constrained the covalent examples 5MAJ (Giroud et al. 2017) and 5F2E (Patricelli et al. 2016). The model then wrote scripts to search ligand positions and orientations, rotate flexible bonds, and test protonation or tautomer states, using RDKit (RDKit n.d.)[2] to generate ligand shapes and MMFF (Merck Molecular Force Field; Halgren 1996) molecular mechanics to reduce internal strain. Custom scoring terms favored hydrogen bonds, metal coordination, and plausible covalent geometry while penalizing overlapping atoms and strained bonds. The model compared competing solutions, including 7BJJ (Williamson et al. 2021) tautomers with the ring hydrogen on N6 or N7 and average structures representing distinct groups of sampled poses in 7UY4.

Refinement incorporated information from multiple candidates: in 7BJJ, the model rebuilt side chains and used a target combining 79% of the ensemble mean with 21% of the leading docking pose. It also addressed specific defects, for example, increasing the separation between Glu190 OE2 and water W3 from 2.18 to 2.44 Å to relieve a crowded contact in 5MAJ. In 5F2E, it flagged an amide nitrogen displaced from its expected plane in a charged result from hybrid quantum/classical refinement; the submitted pose came from a separate neutral branch.

7 Conclusion — Towards Programmable Molecular Design

The GPT-6 model series demonstrates strong intrinsic capabilities in protein biology, with a notable exception in end-to-end structure prediction. Our protein design campaigns still rely on specialized structure prediction models, particularly for filtering candidates. We also acknowledge the decades of experimental work and data curation that produced the biological datasets underpinning this research, including protein structure data.

While lab testing remains a critical part of protein design, the intrinsic capabilities demonstrated here provide a hint of a future where powerful models can shorten each round of design by providing more control and flexibility, and simultaneously reduce the total number of design-make-test cycles by increasing the probability of success.

Our north star is to improve human outcomes by advancing models for programmable drug design. Achieving this goal will require substantial additional progress. Most of our campaigns have focused on binding. Our next priority is to optimize functional activity and potency in relevant cellular and in vivo systems. Depending on the intended mechanism, this may involve receptor blockade, immune effector activity, or internalization. We must also optimize developability and specificity while assessing and reducing immunogenicity risk. These goals will require creative experimental designs and rigorous data collection. Finally, we aim to tackle challenging targets and expand beyond VHHs and scFvs to include further modalities such as enzymes, hormones, cytokines, multispecific antibodies, and macrocycles.

[2]RDKit is included as a minimal coding tool.

8 Methods

8.1 Models and Evaluation Settings

We evaluated GPT-5.6, GPT-6, GPT-6.2 and internal models. In the standard setting for Sections 2–4, GPT could write and run code but could not call pretrained protein models or browse the internet. Where indicated, we used specialized models after submission to assign sequences to generated backbones, predict structures or score designs. The experimental campaigns used a pre-release version of GPT-6.2.

For comparisons across inference budgets, we report the mean number of completion tokens generated per task at each setting. Comparisons at the highest budget use the highest setting evaluated for each model and task. Appendix A gives the task scores and specialized-model references used in the overview.

8.2 Benchmarks and Scoring

Protein inverse folding and monomer folding used experimental structures from the RCSB Protein Data Bank (PDB, Berman et al. 2000) and predicted structures from the AlphaFold Database (AFDB, Varadi et al. 2022). SAbDab (Dunbar et al. 2014) supplied antibody structures for the inverse folding and docking benchmarks.

Sequence recovery is the fraction of evaluated positions that match the reference sequence. To measure structural self-consistency, we refolded sequences with ESMFold2 (Candido et al. 2026) and compared the predicted and reference backbones. For protein inverse folding, we also report TM-score (Zhang and Skolnick 2004).

For backbone generation, we evaluated sequences assigned after submission by SolubleMPNN (Goverde et al. 2024), LigandMPNN (Dauparas et al. 2025) or ProteinMPNN (Dauparas et al. 2022). A backbone passed if any one refold met all the criteria in the table, except in Figure 5A, where we scored only the selected refold.

For MotifBench, we pooled successes across motifs and divided by the number of evaluable scaffolds. For the internal model, a scaffold was evaluable if any refold passed or all eight had complete scores.

For each refold, we calculated two root-mean-square deviations (RMSDs) after separate least-squares alignments. Motif RMSD compares the N, Cα and C atoms of the refolded motif with those of the supplied motif. Scaffold self-consistency RMSD (scRMSD) compares all Cα atoms of the refolded scaffold with those of the submitted backbone. To select the example structures in Figure 8D, we applied the same cutoffs using N, Cα, C and O atoms for both comparisons.

For Figures 3A and 5A, we divided the number of successful backbones by the number of eligible trials for each model and target. In Figure 3A, we weighted directional target-to-binder ipTM by target segment length. Figure 3B reports global ipTM for the selected refold. For protein targets, we measured binder Cα RMSD after aligning the targets and binder backbone RMSD over N, Cα, C and O after aligning the binders.

We also computed ipSAE (Dunbrack 2025) for protein targets. The criteria below follow Pacesa et al. (2025) for protein minibinders and Campbell et al. (2024) for protein inverse folding. Section 8.3 describes selection for experimental testing.

Benchmark	Unit of evaluation	Computational success criteria
Protein minibinders	Backbone, eight SolubleMPNN sequences	Target-aligned RMSD < 5 Å, binder-only RMSD < 3 Å and directional target-to-binder ipTM ≥ 0.5
RNA, DNA and small-molecule minibinders	Backbone, eight LigandMPNN sequences	Global ipTM ≥ 0.5 and RMSD ≤ 5 Å for the selected refold. RMSD is measured over binder Cα atoms after target phosphate alignment for RNA and DNA, or over ligand atoms after binder backbone alignment for small molecules.
CATH-conditioned generation	Backbone, up to eight ProteinMPNN sequences	Requested topology and backbone RMSD < 2 Å after direct alignment
Motif scaffolding	Scaffold, eight ProteinMPNN sequences	Motif RMSD < 1.5 Å and scaffold scRMSD < 2 Å
Protein inverse folding	Target with RMSD available for every required chain	Backbone self-consistency RMSD < 2 Å for every required chain
Antibody inverse folding	Antibody, complete tasks only	Mean self-consistency RMSD across applicable CDRs < 2 Å

8.3 Selection for Experimental Testing

We evaluated each candidate complex with ESMFold2 (Candido et al. 2026), Boltz-2 (Passaro et al. 2025), and OpenDDE (Aureka AI OpenDDE project 2026). For a candidate to be eligible for testing, at least two of the three models had to meet all structural criteria for that campaign. We ranked eligible designs by mean ipSAE (Dunbrack 2025) across all three models, using the minimum directional score for antibodies.

Campaign	Interface criteria	Backbone and sequence criteria
De novo minibinders (Section 2.5)	Minimum ipSAE ≥ 0.50 and ipTM ≥ 0.5	Binder Cα RMSD < 2.2 Å after target alignment
Minibinder inverse folding (Section 3.3.1)	Minimum target-group ipSAE ≥ 0.5	Binder Cα RMSD < 2.2 Å after target alignment
VHH inverse folding (Section 3.3.2)	Minimum directional ipSAE ≥ 0.5	Binder backbone RMSD ≤ 4.0 Å after target alignment and backbone RMSD ≤ 2.0 Å for each heavy-chain CDR after binder alignment
Therapeutic antibody redesign (Section 3.3.3)	Same as VHH inverse folding	Same RMSD criteria as VHH inverse folding, plus substitutions at $> 40\%$ of HCDR3 positions

Computational pass rates in Section 3.3 refer to ESMFold2 alone. Experimental hit rates are the fraction of designs with a reported screening result that were classified as binders. Appendix B describes single-dose biolayer interferometry (BLI) screening and the separate dose-response measurements used to characterize binding affinity.

Acknowledgements

We gratefully acknowledge Laurance Fauconnet, Jyoti Sheldon, Shrikala Kashyap, Chris Butcher, Amelia Liu, Yunyun Wang, Feitong Yang, Daniel Wolf, Maxine Mackintosh, Chris Hayduk, Ashoka Rajendra, David Yang, Ahmed El-Kishky, Edmund Wong, Tongzhou Wang, and the entire team at OpenAI for their support throughout this project. We also thank Ido Calman, Zeynep Onder, and Hamed Shaykhalishahi at Onava for their contributions to our protein design campaign.

References

A-Alpha Bio. 2025. "Open Protein–Protein Interaction Affinity Datasets with AlphaSeq." Hugging Face. https://huggingface.co/datasets/aalphabio/open-alphaseq.

Abramson, J., J. Adler, J. Dunger, et al. 2024. "Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3." Nature 630: 493–500. https://doi.org/10.1038/s41586-024-07487-w.

Adams, Chloe S., Hyojin Kim, Abigail E. Burtner, et al. 2025. "De Novo Design of Protein Minibinder Agonists of TLR3." Nature Communications 16: 1234. https://doi.org/10.1038/s41467-025-56369-w.

Agarwal, Aditya A., James Harrang, David Noble, et al. 2025. "AlphaBind, a Domain-Specific Model to Predict and Optimize Antibody–Antigen Binding Affinity." mAbs 17 (1): 2534626. https://doi.org/10.1080/19420862.2025.2534626.

Antonyuk, Svetlana, Richard W. Strange, and S. Samar Hasnain. 2010. "Structural Discovery of Small Molecule Binding Sites in Cu–Zn Human Superoxide Dismutase Familial Amyotrophic Lateral Sclerosis Mutants Provides Insights for Lead Optimization." Journal of Medicinal Chemistry 53 (3): 1402–6. https://doi.org/10.1021/jm9017948.

Arora, Rohit Krishan, Leo Tianlai Chen, Melissa Du, Debora Marks, and George Church. 2026. "PG-LLM: Benchmarking General-Purpose Language Models for Protein Variant Ranking." Preprint. bioRxiv. https://doi.org/10.64898/2026.07.27.741045.

Aureka AI OpenDDE project. 2026. "Folding, Reasoning, and Scaling with Open-Source Drug Discovery Engine." Preprint 2607.03787. July 4. arXiv. https://doi.org/10.48550/arXiv.2607.03787.

Basu, Sankar, and Björn Wallner. 2016. "DockQ: A Quality Measure for Protein-Protein Docking Models." PLOS ONE 11 (8): e0161879. https://doi.org/10.1371/journal.pone.0161879.

Bentley, G. A., G. Boulot, M. M. Riottot, and R. J. Poljak. 1990. "Three-Dimensional Structure of an Idiotope-Anti-Idiotope Complex." Nature 348 (6298): 254–57. https://doi.org/10.1038/348254a0.

Berisio, R., A. Viguera, L. Serrano, and M. Wilmanns. 2001. "Atomic Resolution Structure of a Mutant of the Spectrin SH3 Domain." Acta Crystallographica Section D: Biological Crystallography 57 (2): 337–40. https://doi.org/10.1107/S090744490001581X.

Berman, Helen M., John Westbrook, Zukang Feng, et al. 2000. "The Protein Data Bank."

Nucleic Acids Research 28 (1): 235–42. https://doi.org/10.1093/nar/28.1.235.

Brown, M., M. A. Schumacher, G. D. Wiens, R. G. Brennan, and M. B. Rittenberg. 2000. "The Structural Basis of Repertoire Shift in an Immune Response to Phosphocholine." Journal of Experimental Medicine 191 (12): 2101–12. https://doi.org/10.1084/jem.191.12.2101.

Butcher, Jasper, Jie Chen, Andrew Hunt, et al. 2026. "Reliable and Efficient Agentic Protein Design with Toolset Harnessing." Preprint. October 20. bioRxiv. https://doi.org/10.64898/2026.10.19.759896.

Butcher, Jasper, Rohith Krishna, Raktim Mitra, et al. 2026. "Atomic Functional Site Scaffolding in Biopolymers with RFdiffusion3." Science, October 15, eaeh2982. https://doi.org/10.1126/science.aeh2982.

Buttenschoen, Martin, Garrett M. Morris, and Charlotte M. Deane. 2024. "PoseBusters: AI-Based Docking Methods Fail to Generate Physically Valid Poses or Generalise to Novel Sequences." Chemical Science 15 (9): 3130–39. https://doi.org/10.1039/D3SC04185A.

Calman, Ido, Hamed Shaykhalishahi, Puwanat Sangkhapreecha, and Sharrol Bachas. 2025. "Iterative Epitope Expansion Enables Structure-Guided de Novo Design of IgE-Binding Miniproteins Targeting the FcϵRI Interface." Preprint. December 27. bioRxiv.https://doi.org/10.64898/2025.12.27.696647

Campbell, Andrew, Jason Yim, Regina Barzilay, Tom Rainforth, and Tommi Jaakkola. 2024. "Generative Flows on Discrete State-Spaces: Enabling Multimodal Flows with Applications to Protein Co-Design." Proceedings of the 41st International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 235: 5453–512. https://proceedings.mlr.press/v235/campbell24a.html.

Candido, Salvatore, Thomas Hayes, Alexander Derry, et al. 2026. "Language Modeling Materializes a World Model of Protein Biology." Preprint. June 4. bioRxiv. https://doi.org/10.64898/2026.06.03.729735.

Cao, Longxing, Brian Coventry, Inna Goreshnik, et al. 2022. "Design of Protein-Binding Proteins from the Target Structure Alone." Nature 605 (7910): 551–60. https://doi.org/10.1038/s41586-022-04654-9.

Chen, Xi, Kathrin Hnida, Melissa Ann Graewert, et al. 2015. "Structural Basis for Antigen Recognition by Transglutaminase 2-Specific Autoantibodies in Celiac Disease." Journal of Biological Chemistry 290 (35): 21365–75. https://doi.org/10.1074/jbc.M115.669895.

Chevalier, Aaron, Daniel-Adriano Silva, Gabriel J. Rocklin, et al. 2017. "Massively Parallel de Novo Protein Design for Targeted Therapeutics." Nature 550 (7674): 74–79. https://doi.org/10.1038/nature23912.

Cho, Yehlin, Martin Pacesa, Zhidian Zhang, Bruno E. Correia, and Sergey Ovchinnikov. 2025. "Boltzdesign1: Inverting All-Atom Structure Prediction Model for Generalized Biomolecular Binder Design." Preprint. April 6. bioRxiv. https://doi.org/10.1101/2025.04.06.647261.

Chothia, Cyrus, and Arthur M. Lesk. 1987. "Canonical Structures for the Hypervariable Regions of Immunoglobulins." Journal of Molecular Biology 196 (4): 901–17. https://doi.org/10.1016/0022-2836(87)90412-8.

Claude Science, and Amir Shanehsazzadeh. 2026. Autonomous de Novo Protein Binder Design with Claude. Technical report. Anthropic. https://assets.claude.com/30bf50e22a01388bb29bf077ee3f244531594b7a.pdf.

Cody, Vivian, Jim Pace, Hesham F. Nawar, et al. 2012. "Structure-Activity Correlations of Variant Forms of the B Pentamer of Escherichia coli Type II Heat-Labile Enterotoxin LT-IIb with Toll-Like Receptor 2 Binding." Acta Crystallographica Section D: Biological Crystallography 68 (12): 1604–12. https://doi.org/10.1107/S0907444912038917.

Corso, Gabriele, Arthur Deng, Benjamin Fry, Nicholas Polizzi, Regina Barzilay, and Tommi Jaakkola. 2024. "Deep Confident Steps to New Pockets: Strategies for Docking Generalization." International Conference on Learning Representations. https://doi.org/10.48550/arXiv.2402.18396.

Corso, Gabriele, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. 2023. "DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking." International Conference on Learning Representations. https://doi.org/10.48550/arXiv.2210.01776.

Craig, Nathaniel. 2026. "A Gauge-Invariant Measure for Lattice Chiral Fermions." Preprint. October 6. arXiv. https://doi.org/10.48550/arXiv.2610.09093.

Dauparas, Justas, I. Anishchenko, N. Bennett, et al. 2022. "Robust Deep Learning–Based Protein Sequence Design Using ProteinMPNN." Science 378 (6615): 49–56. https://doi.org/10.1126/science.add2187.

Dauparas, Justas, Gyu Rie Lee, Robert Pecoraro, et al. 2025. "Atomic Context-Conditioned Protein Sequence Design Using LigandMPNN." Nature Methods 22 (4): 717–23. https://doi.org/10.1038/s41592-025-02626-1.

Didi, Kieran, Sarah Alamdari, Alex Xijie Lu, et al. 2026. "FLIP2: Expanding Protein Fitness Landscape Benchmarks for Real-World Machine Learning Applications." Proceedings of the 43rd International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 306: 24724–77. https://proceedings.mlr.press/v306/didi26a.html.

Dreyer, Frédéric A., Daniel Cutting, Constantin Schneider, Henry Kenlay, and Charlotte M. Deane. 2023. "Inverse Folding for Antibody Sequence Design Using Deep Learning." 2023 ICML Workshop on Computational Biology. https://icml-compbio.github.io/2023/papers/WCBICML2023_paper61.pdf.

Du, Siyuan, Rachael C. Kretsch, Jacob Parres-Gold, et al. 2024. "Conformational Ensembles Reveal the Origins of Serine Protease Catalysis." Preprint 2024.02.28.582624v2. March 5. bioRxiv. https://doi.org/10.1101/2024.02.28.582624.

Dunbar, James, Konrad Krawczyk, Jinwoo Leem, et al. 2014. "SAbDab: The Structural Antibody Database." Nucleic Acids Research 42 (D1): D1140–46. https://doi.org/10.1093/nar/gkt1043.

Dunbrack, Roland L., Jr. 2025. "Rēs ipSAE Loquuntur: What's Wrong with AlphaFold's ipTM Score and How to Fix It." Preprint. February 14. bioRxiv. https://doi.org/10.1101/2025.02.10.637595.

Evans, Richard, Michael O'Neill, Alexander Pritzel, et al. 2021. "Protein Complex Prediction with AlphaFold-Multimer." Preprint. October 4. bioRxiv. https://doi.org/10.1101/2021.10.04.463034.

Fedechkin, Stanislav O., Natasha L. George, Jacob T. Wolff, Lawrence M. Kauvar, and Rebecca M. DuBois. 2018. "Structures of Respiratory Syncytial Virus G Antigen Bound to Broadly Neutralizing Antibodies." Science Immunology 3 (21): eaar3534. https://doi.org/10.1126/sciimmunol.aar3534.

Franklin, Matthew C., Elizabeth C. Navarro, Yujie Wang, et al. 2011. "The Structural Basis for the Function of Two Anti-VEGF Receptor 2 Antibodies." Structure 19 (8): 1097–1107. https://doi.org/10.1016/j.str.2011.01.019.

Frolow, F., A. J. Kalb, and J. Yariv. 1994. "Structure of a Unique Twofold Symmetric Haem-Binding Site." Nature Structural Biology 1 (7): 453–60. https://doi.org/10.1038/nsb0794-453.

Geffner, Tomas, Kieran Didi, Zuobai Zhang, et al. 2025. "Proteína: Scaling Flow-Based Protein Structure Generative Models." International Conference on Learning Representations. https://openreview.net/forum?id=TVQLu34bdw.

Gelman, Sam, Bryce Johnson, Chase R. Freschlin, et al. 2025. "Biophysics-Based Protein Language Models for Protein Engineering." Nature Methods 22 (9): 1868–79. https://doi.org/10.1038/s41592-025-02776-2.

Giroud, Maude, Jakov Ivkovic, Mara Martignoni, et al. 2017. "Inhibition of the Cysteine Protease Human Cathepsin L by Triazine Nitriles: Amide···Heteroarene π-Stacking Interactions and Chalcogen Bonding in the S3 Pocket." ChemMedChem 12 (3): 257–70. https://doi.org/10.1002/cmdc.201600563.

Goverde, Casper A., Martin Pacesa, Nicolas Goldbach, et al. 2024. "Computational Design of Soluble and Functional Membrane Protein Analogues." Nature 631 (8020): 449–58. https://doi.org/10.1038/s41586-024-07601-y.

Greisen, Per Jr., Li Yi, Rong Zhou, et al. 2023. "Computational Design of N-Linked Glycans for High Throughput Epitope Profiling." Protein Science 32 (10): e4726. https://doi.org/10.1002/pro.4726.

Gruszka, Dominika T., Justyna A. Wojdyla, Richard J. Bingham, et al. 2012. "Staphylococcal Biofilm-Forming Protein Has a Contiguous Rod-Like Structure." Proceedings of the National Academy of Sciences 109 (17): E1011–18. https://doi.org/10.1073/pnas.1119456109.

Guillet, V., A. Lapthorn, R. W. Hartley, and Y. Mauguen. 1993. "Recognition between a Bacterial Ribonuclease, Barnase, and Its Natural Inhibitor, Barstar." Structure 1 (3): 165–76. https://doi.org/10.1016/0969-2126(93)90018-C.

Halgren, Thomas A. 1996. "Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94." Journal of Computational Chemistry 17 (5–6): 490–519. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6%3C490::AID-JCC1%3E3.0.CO;2-P.

Hayes, Thomas, Roshan Rao, Halil Akin, et al. 2024. "Simulating 500 Million Years of Evolution with a Language Model." Preprint. July 2. bioRxiv. https://doi.org/10.1101/2024.07.01.600583.

Hong, Yuan, Yingcai Feng, Hanzi Sun, et al. 2021. "Tislelizumab Uniquely Binds to the CC′ Loop of PD-1 with Slow-Dissociated Rate and Complete PD-L1 Blockage." FEBS Open Bio 11 (3): 782–92. https://doi.org/10.1002/2211-5463.13102.

Huang, Kexin, Serena Zhang, Hanchen Wang, et al. 2026. "Autonomous Biomedical Research with an Artificial Intelligence Agent." Science 393 (6813): eadz4351. https://doi.org/10.1126/science.adz4351.

Ingraham, John B., Max Baranov, Zak Costello, et al. 2023. "Illuminating Protein Space with a Programmable Generative Model." Nature 623 (7989): 1070–78. https://doi.org/10.1038/s41586-023-06728-8.

Issafras, Hassan, Shilong Fan, Chi-Ling Tseng, et al. 2021. "Structural Basis of HLX10 PD-1 Receptor Recognition, a Promising Anti-PD-1 Antibody Clinical Candidate for Cancer Immunotherapy." PLOS ONE 16 (12): e0257972. https://doi.org/10.1371/journal.pone.0257972.

Jacobson, Frida, Arthur Pistorius, Daniel Farkas, et al. 2007. "pH Dependence of Copper Geometry, Reduction Potential, and Nitrite Affinity in Nitrite Reductase." Journal of Biological Chemistry 282 (9): 6347–55. https://doi.org/10.1074/jbc.M605746200.

Jiang, Lin, Eric A. Althoff, Fernando R. Clemente, et al. 2008. "De Novo Computational Design of Retro-Aldol Enzymes." Science 319 (5868): 1387–91. https://doi.org/10.1126/science.1152692.

Jiménez-García, Brian, Jorge Roel-Touris, Miguel Romero-Durana, Miquel Vidal, Daniel Jiménez-González, and Juan Fernández-Recio. 2018. "LightDock: A New Multi-Scale Approach to Protein–Protein Docking." Bioinformatics 34 (1): 49–55. https://doi.org/10.1093/bioinformatics/btx555.

Jumper, John, Richard Evans, Alexander Pritzel, et al. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold." Nature 596 (7873): 583–89. https://doi.org/10.1038/s41586-021-03819-2.

Kabsch, Wolfgang. 1976. "A Solution for the Best Rotation to Relate Two Sets of Vectors." Acta Crystallographica Section A 32 (5): 922–23. https://doi.org/10.1107/S0567739476001873.

Kandathil, Shaun M., Andy M. Lau, Daniel W. A. Buchan, and David T. Jones. 2025. "Foldclass and Merizo-Search: Scalable Structural Similarity Search for Single- and Multi-Domain Proteins Using Geometric Learning." Bioinformatics 41 (5): btaf277. https://doi.org/10.1093/bioinformatics/btaf277.

Köster, Stefan, Mark Wehner, Christian Herrmann, Werner Kühlbrandt, and Ozkan Yildiz. 2009. "Structure and Function of the FeoB G-Domain from Methanococcus jannaschii." Journal of Molecular Biology 392 (2): 405–19. https://doi.org/10.1016/J.JMB.2009.07.020.

Kovalev, Kirill, Vitaly Polovinkin, Ivan Gushchin, et al. 2019. "Structure and Mechanisms of Sodium-Pumping KR2 Rhodopsin." Science Advances 5 (4): eaav2671. https://doi.org/10.1126/sciadv.aav2671.

Krasnow, Nicholas A., Joy A. Xu, Emily Zhang, et al. 2026. "AI-Redesigned Starting Points and Outcomes Enhance Protein Evolution." Nature 657: 525–538. https://doi.org/10.1038/s41586-026-10820-0.

Lau, Andy M., Shaun M. Kandathil, and David T. Jones. 2023. "Merizo: A Rapid and Accurate Protein Domain Segmentation Method Using Invariant Point Attention." Nature Communications 14 (1): 8445. https://doi.org/10.1038/s41467-023-43934-4.

Lee, Jin Sub, and Philip M. Kim. 2025. "FlowPacker: Protein Side-Chain Packing with Torsional Flow Matching." Bioinformatics 41 (3): btaf010. https://doi.org/10.1093/bioinformatics/btaf010.

Lieu, Ricky, Stephen Antonysamy, Zhanna Druzina, et al. 2020. "Rapid and Robust Antibody Fab Fragment Crystallization Utilizing Edge-to-Edge Beta-Sheet Packing." PLOS ONE 15 (9): e0232311. https://doi.org/10.1371/journal.pone.0232311.

Lin, Zeming, Halil Akin, Roshan Rao, et al. 2023. "Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model." Science 379 (6637): 1123–30. https://doi.org/10.1126/science.ade2574.

Liu, Kefang, Shuguang Tan, Yan Chai, et al. 2017. "Structural Basis of Anti-PD-L1 Monoclonal Antibody Avelumab for Tumor Therapy." Cell Research 27 (1): 151–53. https://doi.org/10.1038/cr.2016.102.

Lu, Wei, Jixian Zhang, Ming Gu, and Shuangjia Zheng. 2024. "BindingGYM: A Large-Scale Mutational Dataset Toward Deciphering Protein-Protein Interactions." AI for New Drug Modalities at NeurIPS. https://openreview.net/forum?id=ZJkqmSgZeH.

Lu, Zhibing, Debra Dunaway-Mariano, and Karen N. Allen. 2008. "The Catalytic Scaffold of the Haloalkanoic Acid Dehalogenase Enzyme Superfamily Acts as a Mold for the Trigonal Bipyramidal Transition State." Proceedings of the National Academy of Sciences 105 (15): 5687–92. https://doi.org/10.1073/pnas.0710800105.

Mariani, Valerio, Marco Biasini, Alessandro Barbato, and Torsten Schwede. 2013. "lDDT: A Local Superposition-Free Score for Comparing Protein Structures and Models Using Distance Difference Tests." Bioinformatics 29 (21): 2722–28. https://doi.org/10.1093/bioinformatics/btt473.

Masters, Matthew R., Amr H. Mahmoud, and Markus A. Lill. 2025. "Investigating Whether Deep Learning Models for Co-Folding Learn the Physics of Protein-Ligand Interactions." Nature Communications 16: 8854. https://doi.org/10.1038/s41467-025-63947-5.

Meier, Joshua, Roshan Rao, Robert Verkuil, Jason Liu, Tom Sercu, and Alex Rives. 2021. "Language Models Enable Zero-Shot Prediction of the Effects of Mutations on Protein Function." Advances in Neural Information Processing Systems 34: 29287–303. https://papers.neurips.cc/paper_files/paper/2021/hash/f51338d736f95dd42427296047067694-Abstract.html.

Mikol, V. 1996. "Structure of the Fab Fragment of SDZ CHI621: A Chimeric Antibody against CD25." Acta Crystallographica Section D: Biological Crystallography 52 (3): 534–42. https://doi.org/10.1107/S0907444996000704.

Mirabello, Claudio, and Björn Wallner. 2024. "DockQ v2: Improved Automatic Quality Measure for Protein Multimers, Nucleic Acids, and Small Molecules." Bioinformatics 40 (10): btae586. https://doi.org/10.1093/bioinformatics/btae586.

Mylemans, Bram, Hiroki Noguchi, Els Deridder, Eveline Lescrinier, Jeremy R. H. Tame, and Arnout R. D. Voet. 2020. "Influence of Circular Permutations on the Structure and Stability of a Six-Fold Circular Symmetric Designer Protein." Protein Science 29 (12): 2375–86. https://doi.org/10.1002/pro.3961.

Na, Zhenkun, Siok Ping Yeo, Sakshibeedu R. Bharath, et al. 2017. "Structural Basis for Blocking

PD-1-Mediated Immune Suppression by Therapeutic Antibody Pembrolizumab." Cell Research 27 (1): 147–50. https://doi.org/10.1038/cr.2016.77.

Notin, Pascal, Aaron Kollasch, Daniel Ritter, et al. 2023. "ProteinGym: Large-Scale Benchmarks for Protein Fitness Prediction and Design." Advances in Neural Information Processing Systems 36: 64331–79. https://doi.org/10.52202/075280-2810.

OpenAI. 2026. "GPT-6 Astra: A New Generation of Intelligence." OpenAI. https://openai.com/index/gpt-6-astra/.

Orengo, C. A., A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton. 1997. "CATH – A Hierarchic Classification of Protein Domain Structures." Structure 5 (8): 1093–108. https://doi.org/10.1016/s0969-2126(97)00260-8.

Pacesa, Martin, Lennart Nickel, Christian Schellhaas, et al. 2025. "One-Shot Design of Functional Protein Binders with BindCraft." Nature 646 (8084): 483–92. https://doi.org/10.1038/s41586-025-09429-6.

Parkkinen, Tarja, Tarja K. Nevanen, Anu Koivula, and Juha Rouvinen. 2006. "Crystal Structures of an Enantioselective Fab-Fragment in Free and Complex Forms." Journal of Molecular Biology 357 (2): 471–80. https://doi.org/10.1016/J.JMB.2005.12.045.

Passaro, Saro, Gabriele Corso, Jeremy Wohlwend, et al. 2025. "Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction." Preprint. June 18. bioRxiv. https://doi.org/10.1101/2025.06.14.659707.

Patricelli, Matthew P., Matthew R. Janes, Lian-Sheng Li, et al. 2016. "Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State." Cancer Discovery 6 (3): 316–29. https://doi.org/10.1158/2159-8290.CD-15-1105.

Pichla, S. L., R. Murali, and R. M. Burnett. 1997. "The Crystal Structure of a Fab Fragment to the Melanoma-Associated GD2 Ganglioside." Journal of Structural Biology 119 (1): 6–16. https://doi.org/10.1006/jsbi.1997.3857.

RDKit. n.d. "RDKit: Open-Source Cheminformatics." Accessed November 24, 2026. https://www.rdkit.org.

Read, J. A. 2020. "SYK Kinase Domain in Complex with Diamine Inhibitor 16." Protein Data Bank, PDB ID 6ZCY. https://doi.org/10.2210/pdb6ZCY/pdb.

Reeve, S. M., R. E. Lee, N. Das, and S. Jayaraman. 2022. "Aminoglycoside-Modifying Enzyme ANT-3,9 in Complex with Spectinomycin and AMP-PNP." Protein Data Bank, PDB ID 7UY4. https://doi.org/10.2210/pdb7UY4/pdb.

Ribeiro, António J. M., Gemma L. Holliday, Nicholas Furnham, Jonathan D. Tyzack, Katherine Ferris, and Janet M. Thornton. 2018. "Mechanism and Catalytic Site Atlas (M-CSA): A Database of Enzyme Reaction Mechanisms and Active Sites." *Nucleic Acids Research* 46 (D1): D618–23. https://doi.org/10.1093/nar/gkx1012.

Rives, Alexander, Joshua Meier, Tom Sercu, et al. 2021. "Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences." Proceedings of the National Academy of Sciences 118 (15): e2016239118. https://doi.org/10.1073/pnas.2016239118.

Röthlisberger, Daniela, Olga Khersonsky, Andrew M. Wollacott, et al. 2008. "Kemp Elimination Catalysts by Computational Enzyme Design." Nature 453 (7192): 190–95. https://doi.org/10.1038/nature06879.

Rotkiewicz, Piotr, and Jeffrey Skolnick. 2008. "Fast Procedure for Reconstruction of Full-Atom Protein Models from Reduced Representations." Journal of Computational Chemistry 29 (9): 1460–65. https://doi.org/10.1002/jcc.20906.

Sathiyamoorthy, Karthik, Yao Xiong Hu, Britta S. Möhl, Jia Chen, Richard Longnecker, and Theodore S. Jardetzky. 2016. "Structural Basis for Epstein-Barr Virus Host Cell Tropism Mediated by gp42 and gHgL Entry Glycoproteins." Nature Communications 7: 13557. https://doi.org/10.1038/ncomms13557.

Scarpino, Andrea, György G. Ferenczy, and György M. Keserű. 2018. "Comparative Evaluation of Covalent Docking Tools." Journal of Chemical Information and Modeling 58 (7): 1441–58. https://doi.org/10.1021/acs.jcim.8b00228.

Schafer, Joseph W., and Lauren L. Porter. 2025. "AlphaFold2's Training Set Powers Its Predictions of Some Fold-Switched Conformations." Protein Science 34 (4): e70105. https://doi.org/10.1002/pro.70105.

Schiweck, W., and A. Skerra. 1997. "The Rational Construction of an Antibody against Cystatin: Lessons from the Crystal Structure of an Artificial Fab Fragment." Journal of Molecular Biology 268 (5): 934–51. https://doi.org/10.1006/jmbi.1997.1006.

Sesterhenn, Fabian, Che Yang, Jaume Bonet, et al. 2020. "De Novo Protein Design Enables the Precise Induction of RSV-Neutralizing Antibodies." Science 368 (6492): eaay5051. https://doi.org/10.1126/science.aay5051.

Shanker, Varun R., Theodora U. J. Bruun, Brian L. Hie, and Peter S. Kim. 2024. "Unsupervised Evolution of Protein and Antibody Complexes with a Structure-Informed Language Model." Science 385 (6704): 46–53. https://doi.org/10.1126/science.adk8946.

Shao, Yaming, Hao Huang, Daoming Qin, et al. 2016. "Specific Recognition of a Single-Stranded RNA Sequence by a Synthetic Antibody Fragment." Journal of Molecular Biology 428 (20): 4100–4114. https://doi.org/10.1016/j.jmb.2016.08.029.

Škrinjar, Peter, Jérôme Eberhardt, Gabriel Studer, Gerardo Tauriello, Torsten Schwede, and Janani Durairaj. 2026. "Evaluating Generalization in Protein–Ligand Cofolding Methods." Nature Structural & Molecular Biology 33 (5): 782–94. https://doi.org/10.1038/s41594-026-01797-5.

Stark, Hannes, Felix Faltings, MinGyu Choi, et al. 2025. "BoltzGen: Toward Universal Binder Design." Preprint. November 24. bioRxiv. https://doi.org/10.1101/2025.11.20.689494.

Sutton, Kristin A., Jennifer Breen, Ulrike MacDonald, et al. 2015. "Structure of Shikimate Kinase, an In Vivo Essential Metabolic Enzyme in the Nosocomial Pathogen Acinetobacter baumannii, in Complex with Shikimate." Acta Crystallographica Section D: Biological Crystallography 71 (8): 1736–44. https://doi.org/10.1107/S139900471501189X.

Sutton, Rich. 2019. "The Bitter Lesson." Incomplete Ideas, March 13. http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

Swanson, Kyle, Wesley Wu, Nash L. Bulaong, John E. Pak, and James Zou. 2025. "The Virtual Lab of AI Agents Designs New SARS-CoV-2 Nanobodies." Nature 646 (8085): 716–23. https://doi.org/10.1038/s41586-025-09442-9.

Tanaka, Nobutada, Masayuki Nakanishi, Yoshio Kusakabe, Yoshikuni Goto, Yukio Kitade, and Kazuo T. Nakamura. 2004. "Structural Basis for Recognition of 2′,5′-Linked Oligoadenylates by Human Ribonuclease L." EMBO Journal 23 (20): 3929–38. https://doi.org/10.1038/sj.emboj.7600420.

Tsuboyama, Kotaro, Justas Dauparas, Jonathan Chen, et al. 2023. "Mega-Scale Experimental Analysis of Protein Folding Stability in Biology and Design." Nature 620 (7973): 434–44. https://doi.org/10.1038/s41586-023-06328-6.

Tsuchiya, Yuko, and Kenji Mizuguchi. 2016. "The Diversity of H3 Loops Determines the Antigen-Binding Tendencies of Antibody CDR Loops." Protein Science 25 (4): 815–25. https://doi.org/10.1002/pro.2874.

UniProt Consortium. 2025. "UniProt: The Universal Protein Knowledgebase in 2025." Nucleic Acids Research 53 (D1): D609–17. https://doi.org/10.1093/nar/gkae1010.

Varadi, Mihaly, Stephen Anyango, Mandar Deshpande, et al. 2022. "AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models." Nucleic Acids Research 50 (D1): D439–44. https://doi.org/10.1093/nar/gkab1061.

Wang, Dacheng, Wolfram Bode, and Robert Huber. 1985. "Bovine Chymotrypsinogen A: X-Ray Crystal Structure Analysis and Refinement of a New Crystal Form at 1.8 Å Resolution." Journal of Molecular Biology 185 (3): 595–624. https://doi.org/10.1016/0022-2836(85)90074-9.

Watson, Joseph L., David Juergens, Nathaniel R. Bennett, et al. 2023. "De Novo Design of Protein Structure and Function with RFdiffusion." Nature 620 (7976): 1089–100. https://doi.org/10.1038/s41586-023-06415-8.

Williamson, Douglas S., Garrick P. Smith, Gitte K. Mikkelsen, et al. 2021. "Design and Synthesis of Pyrrolo[2,3-d]pyrimidine-Derived Leucine-Rich Repeat Kinase 2 (LRRK2) Inhibitors Using a Checkpoint Kinase 1 (CHK1)-Derived Crystallographic Surrogate." Journal of Medicinal Chemistry 64 (14): 10312–32. https://doi.org/10.1021/acs.jmedchem.1c00720.

Wilson, Steven C., Nathanael A. Caveney, Michelle Yen, et al. 2022. "Organizing Structural Principles of the IL-17 Ligand–Receptor Axis." Nature 609 (7927): 622–29. https://doi.org/10.1038/s41586-022-05116-y.

Wu, Ruidong, Fan Ding, Rui Wang, et al. 2022. "High-Resolution de Novo Structure Prediction from Primary Sequence." Preprint. July 22. bioRxiv. https://doi.org/10.1101/2022.07.21.500999.

Xu, Jinrui, and Yang Zhang. 2010. "How Significant Is a Protein Structure Similarity with TM-Score = 0.5?" Bioinformatics 26 (7): 889–95. https://doi.org/10.1093/bioinformatics/btq066.

Yang, Che, Fabian Sesterhenn, Jaume Bonet, et al. 2021. "Bottom-up de Novo Design of Functional Proteins with Complex Structural Features." Nature Chemical Biology 17 (4):

492–500. https://doi.org/10.1038/s41589-020-00699-x.

Ye, Shijie, Bernhard Loll, Allison Ann Berger, et al. 2015. “Fluorine Teams Up with Water to Restore Inhibitor Activity to Mutant BPTI.” Chemical Science 6 (9): 5246–54. https://doi.org/10.1039/c4sc03227f.

Yin, Rui, Brandon Y. Feng, Amitabh Varshney, and Brian G. Pierce. 2022. “Benchmarking AlphaFold for Protein Complex Modeling Reveals Accuracy Determinants.” Protein Science 31 (8): e4379. https://doi.org/10.1002/pro.4379.

Younger, David, Stephanie Berger, David Baker, and Eric Klavins. 2017. “High-Throughput Characterization of Protein–Protein Interactions by Reprogramming Yeast Mating.” Proceedings of the National Academy of Sciences 114 (46): 12166–71. https://doi.org/10.1073/pnas.1705867114.

Zhang, Yang, and Jeffrey Skolnick. 2004. “Scoring Function for Automated Assessment of Protein Structure Template Quality.” Proteins: Structure, Function, and Bioinformatics 57 (4): 702–10. https://doi.org/10.1002/prot.20264.

Zheng, Zhuoqi, Bo Zhang, Kieran Didi, et al. 2025. “MotifBench: A Standardized Protein Design Benchmark for Motif-Scaffolding Problems.” Preprint 2502.12479v2. February 19. arXiv. https://doi.org/10.48550/arXiv.2502.12479.

Appendix

A Details of Figure 2

This section describes the task and evaluation metric for each panel of Figure 2, along with the specialized-model reference where shown.

A. Highest-budget comparison across 14 tasks. The radar groups tasks by category. Experimental-backbone and AlphaFold-backbone inverse folding differ in the source of the input structures. For antibody inverse folding, the radar reports success defined as mean applicable CDR RMSD < 2 Å, and panel H reports scaled round-trip RMSD. For protein-target minibinder design, the radar reports selected global ipTM, and panel C reports selected target-group minimum ipSAE.

B. Motif scaffolding. Performance is measured using an equal-weight composite of scaled motif RMSD and scaled Cα self-consistency RMSD. The specialized-model reference is RFdiffusion.

C. Minibinder design against protein targets. Performance is measured using selected target-group minimum ipSAE. The specialized-model reference is RFdiffusion.

D. Topology-conditioned backbone generation. Performance is measured using an equal-weight composite of CATH hierarchy agreement and scaled round-trip quality. The specialized-model reference is Proteina.

E. Minibinder design against DNA, RNA and small molecules. Performance is measured using selected global ipTM. The specialized-model reference is RFdiffusion3.

F. Inverse folding from experimental backbones. Performance is measured using scaled round-trip RMSD. The specialized-model reference is ProteinMPNN.

G. Inverse folding from AlphaFold-predicted backbones. Performance is measured using scaled round-trip RMSD. The specialized-model reference is ProteinMPNN.

H. Antibody inverse folding. Performance is measured using scaled round-trip RMSD. The specialized-model reference is AbMPNN.

I. Monomer structure prediction on experimental structures (RCSB). Performance is measured using an equal-weight composite of masked backbone lDDT and globally aligned Cα TM-score. The specialized-model reference is ESMFold2.

J. Monomer structure prediction on AlphaFold structures (AFDB). Performance is measured using an equal-weight composite of masked backbone lDDT and globally aligned Cα TM-score. The specialized-model reference is ESMFold2.

K. Antibody CDR structure prediction. Performance is measured using mean loop quality. The specialized-model reference is ESMFold2.

L. Side-chain packing. Performance is measured using scaled atom RMSD. The specialized-model reference is FlowPacker-4.

M. Multimer docking and packing. Performance is measured using GlobalDockQ.

N. Antibody–antigen docking and packing. Performance is measured using grouped

antibody–antigen DockQ.

O. Blind small-molecule docking on PoseBusters. Performance is measured using scaled ligand RMSD. The specialized-model reference is DiffDock-L.

P. Blind small-molecule docking on SCARPiNo. Performance is measured using scaled ligand RMSD. The specialized-model reference is DiffDock-L.

Q. Open AlphaSeq: protein binding strength. Performance is measured using Spearman rank correlation with lab measurements. The specialized-model reference is ESM2.

R. BindingGYM: effects of sequence changes on protein interactions. Performance is measured using Spearman rank correlation with lab measurements. The specialized-model reference is ESM2.

S. ProteinGym: activity, binding, stability and other protein traits. Performance is measured using Spearman rank correlation with lab measurements. The specialized-model reference is ESM2.

T. Mega-Scale (Tsuboyama): folding stability at room temperature. Performance is measured using Spearman rank correlation with lab measurements. This is not a heat-resistance assay. The specialized-model reference is ESM2.

B Experimental Methods

This appendix describes the preparation and experimental characterization of binder designs, from DNA construct design and cell-free expression through single-dose biolayer interferometry (BLI) screening and dose-response validation. It also explains how matched affinity measurements were processed and analyzed to quantify agreement between the two workflows and evaluate whether single-dose apparent affinity estimates can support hit prioritization.

B.1 DNA Construct Design

GPT-generated binders selected for experimental testing were codon optimized for cell-free protein expression and formatted as plate-compatible expression constructs using Onava's proprietary codon optimization and DNA preparation workflow. Each construct encoded the designed binder in the assay format used for its campaign, including single-domain miniprotein, VHH, and scFv formats. DNA constructs were synthesized by Twist Bioscience and then prepared for cell-free protein expression and experimental validation.

Design identities were tracked through synthesis, BLI loading, and dose-response follow-up using exact construct identifiers, short loading sequence IDs, or exact sequence hashes. This linkage connected single-dose measurements to their cognate validation data.

B.2 Single-Dose BLI Screening

Initial binding measurements used Onava's high-throughput BLI workflow, as previously published (Calman et al. 2025). Codon-optimized design constructs were expressed in a cell-free protein

synthesis system and processed in parallel in a multiwell plate format. Expressed designs were immobilized on Gator BLI biosensors and challenged with the cognate soluble antigen at a fixed concentration, typically 1 μM, to obtain a single association and dissociation trajectory for each binder.

Apparent single-dose K_D values were obtained by fitting the sensorgrams with a 1:1 binding model. Each experiment included control binders with known dose-response affinities to calibrate interpretation of the apparent single-dose affinities. Inverse-folding experiments also included the parental scaffold on the same plate as an internal relative-affinity reference.

B.3 Dose-Response BLI Validation

Selected hits from single-dose screening were advanced to dose-response biolayer interferometry (BLI) using the biosensor capture chemistry and molecular format appropriate for each campaign. Biosensors were loaded to a predefined target capture level and challenged with 2-fold serial dilutions of the cognate antigen, starting at 1 μM. Association and dissociation phases were analyzed using a 1:1 binding model, where appropriate, to determine apparent equilibrium dissociation constants (K_D). Validation experiments were performed in duplicate when possible. For lower-affinity binders exhibiting rapid association and dissociation kinetics, association and dissociation times were shortened to 60 seconds each to better resolve the primary binding interaction and minimize secondary binding effects, including time-dependent signal drift and heterogeneous surface interactions. Sensorgrams were aligned to the start of the association phase, and association and dissociation windows were cropped to exclude non-binding transitions and minimize contributions from nonideal signal behavior. Reference subtraction, baseline alignment, kinetic fitting, and data analysis were performed using Gator Bio analysis software. Designs exhibiting substantial deviations from a 1:1 binding model were flagged as atypical, and apparent K_D values were reported only when the sensorgrams supported reliable kinetic fitting.

B.4 Matched Single-Dose and Dose-Response Analysis

Matched comparisons were assembled from design hits with a finite single-dose apparent K_D and at least one finite cognate dose-response K_D. When multiple dose-response replicates were available, finite K_D values were combined as a geometric mean, and the individual replicate values were retained to define the plotted range.

K_D values at the BLI slow-dissociation fit floor, including literal below-floor exports and sub-50 pM finite sentinel values, were capped at 50 pM before conversion to pK_D. Rows with measured single-dose response or Rmax below 0.4 nm were excluded before blinding. Single-dose and dose-response affinities were converted to $pK_D = -\log_{10}(K_D \text{ in M})$, and Pearson and Spearman correlations were computed on the plotted pK_D values.

B.5 Single-Dose BLI Provides Predictive Affinity Estimates for Dose-Response Validation

Each design campaign produced tens to hundreds of candidates, so we used single-dose BLI as the primary quantitative screen before selecting a smaller set of molecules for full titration. Designs

were measured at a fixed analyte concentration (1 μM) and fit to obtain an apparent single-dose K_D. These values provided rank-ordered, assay-normalized estimates of whether each design bound in the expected affinity regime; they were not treated as replacements for full equilibrium dose-response curves. We therefore assessed whether an apparent K_D estimated at one concentration remained predictive when the same binder was subsequently measured in a conventional dose-response experiment.

Across 289 matched design-hit measurements from independent design campaigns and antigen targets, single-dose apparent pK_D values tracked the corresponding dose-response pK_D values with Pearson $r = 0.61$ and Spearman $r_s = 0.66$ (Figure A1). The median absolute error was 3.04-fold, and 222 of 289 measurements fell within 10-fold of the follow-up dose-response value. Although individual one-concentration fits were expected to be approximate, they reproducibly placed binders in the correct affinity range and provided quantitatively useful estimates for high-throughput hit identification.

When relative affinities were measured, control binders with known dose-response affinities calibrated interpretation of the single-dose readout within each assay. In inverse-folding experiments, the parental scaffold provided an internal reference for comparing redesigned variants with the starting binder on the same plate. These calibration points supported hit calling and affinity ranking before follow-up dose-response titration of prioritized designs.

C Generation and Experimental Validation of Parental VHH Binders

To establish starting points for antibody inverse folding, we generated de novo VHH binders targeting human IL-4Rα using an ESM-based gradient-descent generative workflow (Candido et al. 2026). The workflow was orchestrated by GPT-6 through a natural language prompt, with guidance from a human expert. The campaign reused the compact epitope from a productive IL-4Rα miniprotein screen and generated 3,000 ESMC Binder VHH designs against 6WGL chain C (Lieu et al. 2020). A top-300 subset was advanced to paired Boltz-2 and ESMFold2 complex prediction. Of the 287 candidates with complete paired structure predictions, 92 VHHs were selected for synthesis using a combined filter on cross-model complex RMSD, Boltz-2 ipSAE, and ESMFold2 ipSAE.

The 92-design VHH panel was expressed in a cell-free format and screened against human IL-4Rα by single-dose Gator BLI at 1 μM antigen. The BLI analysis platform classified 11 VHHs as binders, corresponding to an 11/92 hit rate (12.0%). Six of these 11 hits passed a conservative apparent-affinity gate requiring a numeric K_D, Full $R^2 \geq 0.80$, and no fitted K_D or dissociation-rate floor.

All 11 platform-classified binders were then retested in duplicate seven-dose BLI titrations, and all retained dose-dependent IL-4Rα binding. Three VHHs exhibited clean, well-resolved kinetics at a 250 nM top antigen concentration. Their replicate apparent K_D values were 13.2 and 9.96 nM for Parent VHH 1, 10.3 and 7.88 nM for Parent VHH 2, and 17.0 and 9.70 nM for Parent VHH 3, spanning 7.88 to 17.0 nM overall (Figure A2). The remaining eight binders reproduced but showed atypical kinetic shapes at the 1 μM top-dose condition. Their exported single-dose and follow-up K_D values were therefore used for engineering triage rather than as final parental affinities.

The sequences and Boltz-2/ESMFold2-predicted target–VHH complex structures of the three ex-

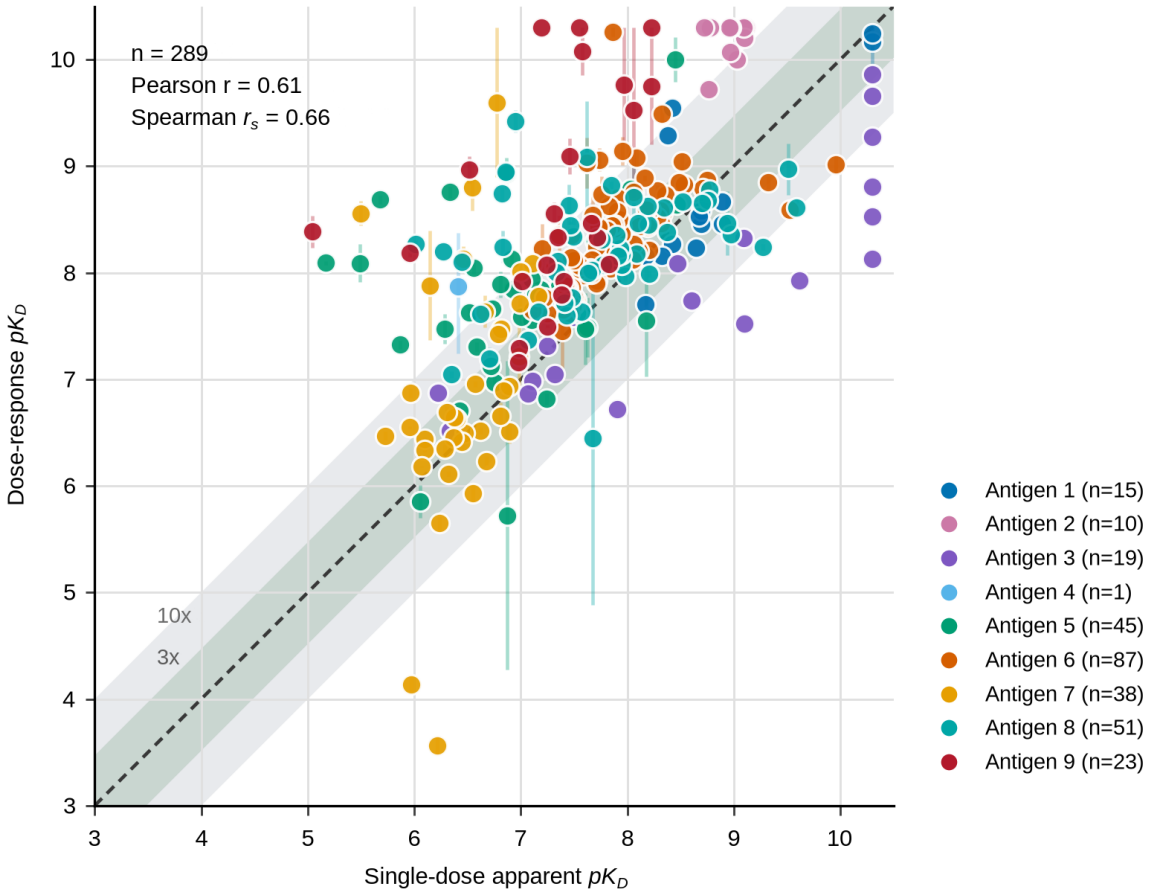


Figure A1. Single-dose BLI apparent affinities agree with follow-up dose-response affinities across multiple antigens. The scatter plot shows 289 matched design-hit comparisons across nine blinded antigens. Each point represents one binder tested in both workflows, and colors denote blinded antigen identity. Single-domain, scFv, Fab, Fc-fusion, and IgG/mAb measurements were included when the same construct or a validated format-conversion match could be linked between the single-dose screen and the cognate dose-response assay. The x and y axes show single-dose apparent pK_D and follow-up dose-response pK_D, respectively, where $pK_D = -\log_{10}(K_D \text{ in M})$. For designs with multiple finite dose-response fits, the plotted y value is the geometric mean K_D converted to pK_D; vertical whiskers span the finite replicate or concentration-specific follow-up K_D values after the same preprocessing. K_D values at the BLI slow-dissociation fit floor were capped at 50 pM before pK_D conversion. Single-dose rows with measured response/Rmax < 0.4 nm were excluded to avoid plotting very low-amplitude traces with artificially fitted K_D values. The dashed diagonal marks perfect agreement; green and gray shading indicate 3-fold and 10-fold agreement, respectively. Correlations on the plotted pK_D values after blinding and filtering were Pearson $r = 0.61$ and Spearman $r_s = 0.66$.

perimentally validated, well-resolved binders were advanced as parental templates for subsequent GPT-6.2-mediated CDR redesign and inverse folding.

D Additional Analyses of Experimental Studies

This appendix summarizes CDR sequence variation across nine VHH families and therapeutic antibody redesigns. It also presents BLI screening traces for parental antibodies and selected

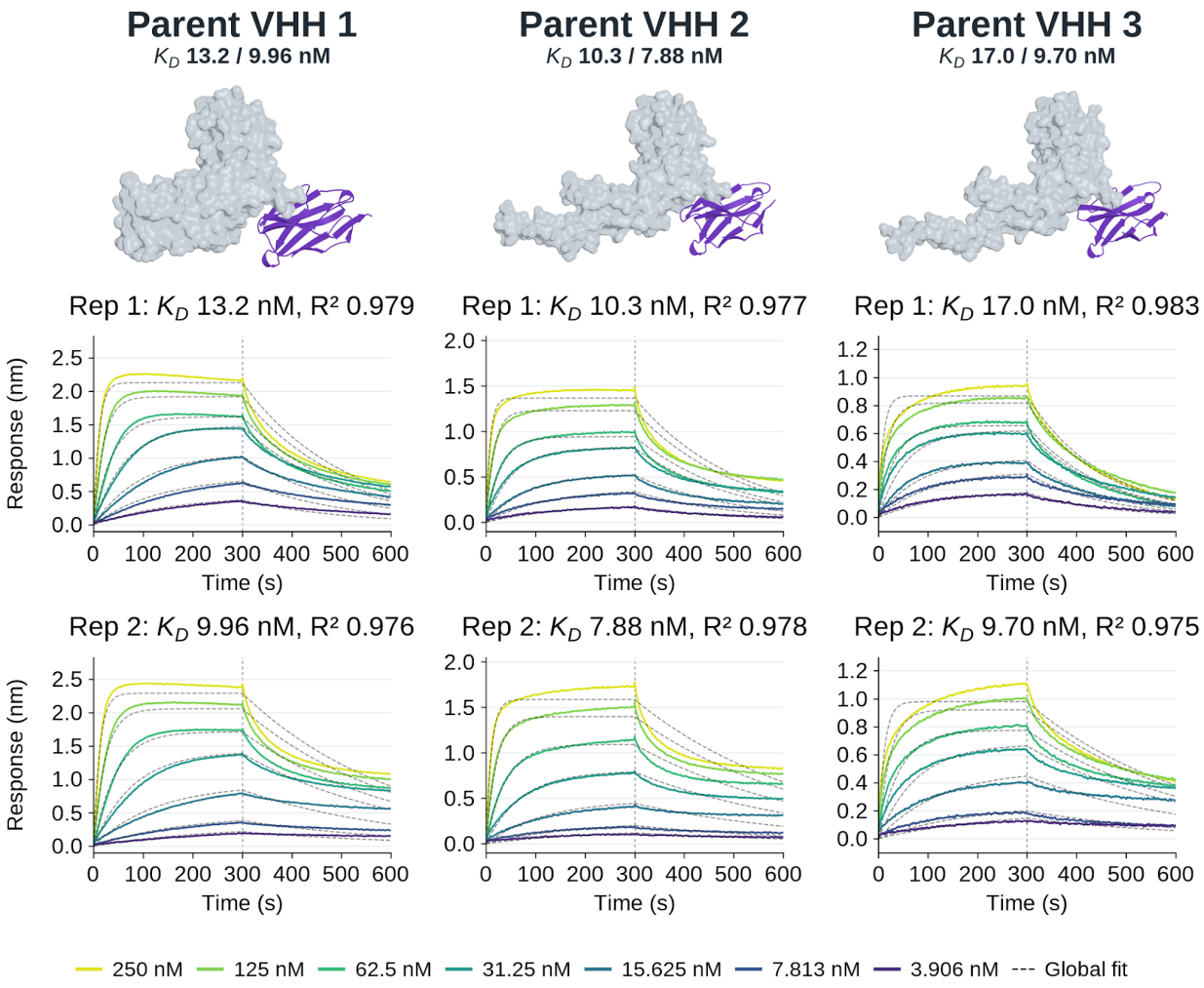


Figure A2. Predicted complex structures and dose-response BLI validation of three parental VHH binders targeting human IL-4Rα. Columns show Parent VHH 1, Parent VHH 2, and Parent VHH 3. The top row shows predicted IL-4Rα–VHH complex structures, with IL-4Rα in gray and VHHs in purple. The middle and bottom rows show two replicate seven-dose BLI titrations per parent, using antigen concentrations from 250 to 3.906 nM. Colored curves show measured responses at the indicated concentrations, and dashed lines show the global fits. Each replicate is labeled with its apparent K_D and R^2.

redesigns.

D.1 CDR Sequence Variation in VHH Inverse Folding

These figures show position-level CDR variation across nine VHH families.

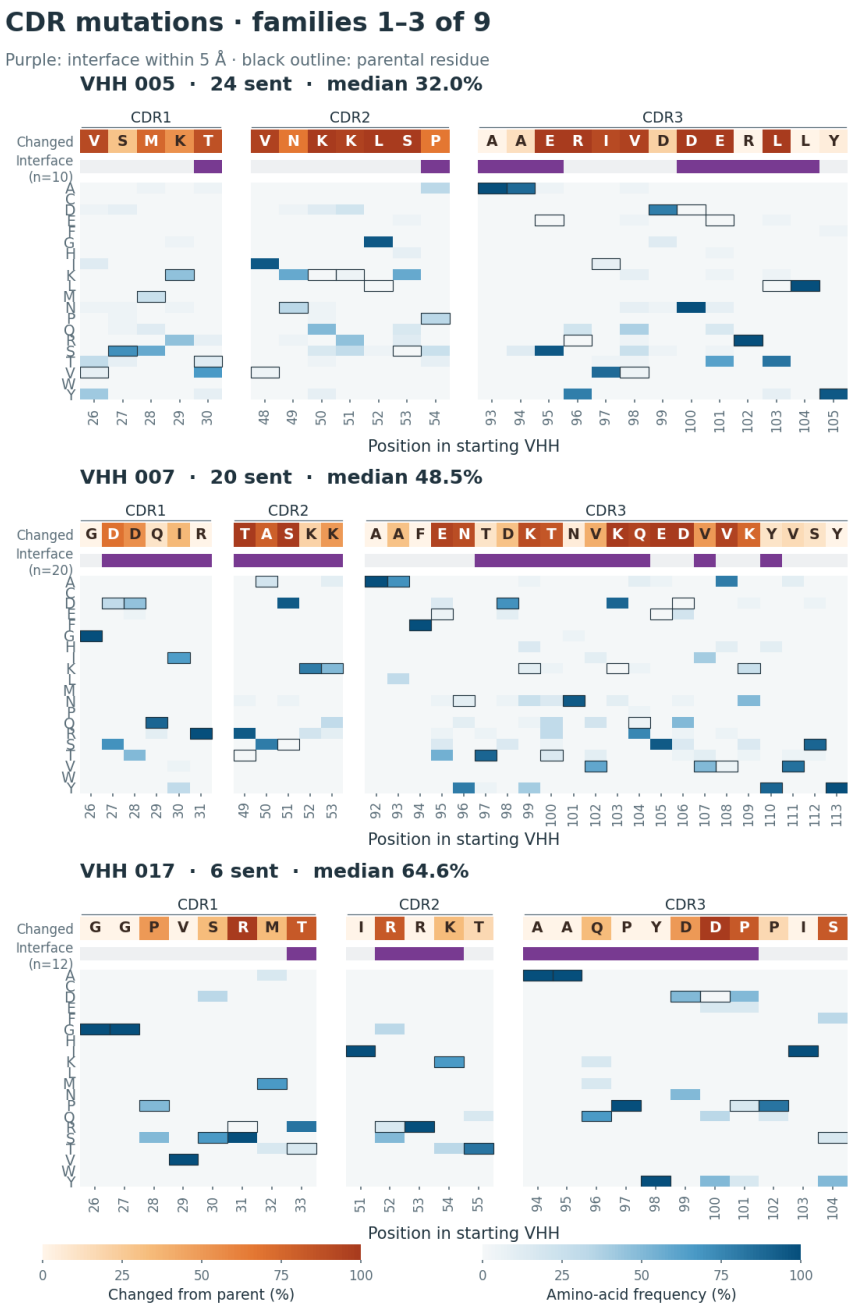


Figure A3A. CDR sequence variation in VHH families 005, 007, and 017. Orange tracks show the percentage of designs changed from the parental residue at each position, and blue heatmaps show amino acid frequencies across the designs sent for testing. Purple tracks mark interface positions within 5 Å of the target, and black outlines identify the parental amino acid.

D.2 BLI Screening Traces for Therapeutic Antibody Redesigns

D.3 HCDR Sequence Variation in Therapeutic Antibody Redesigns

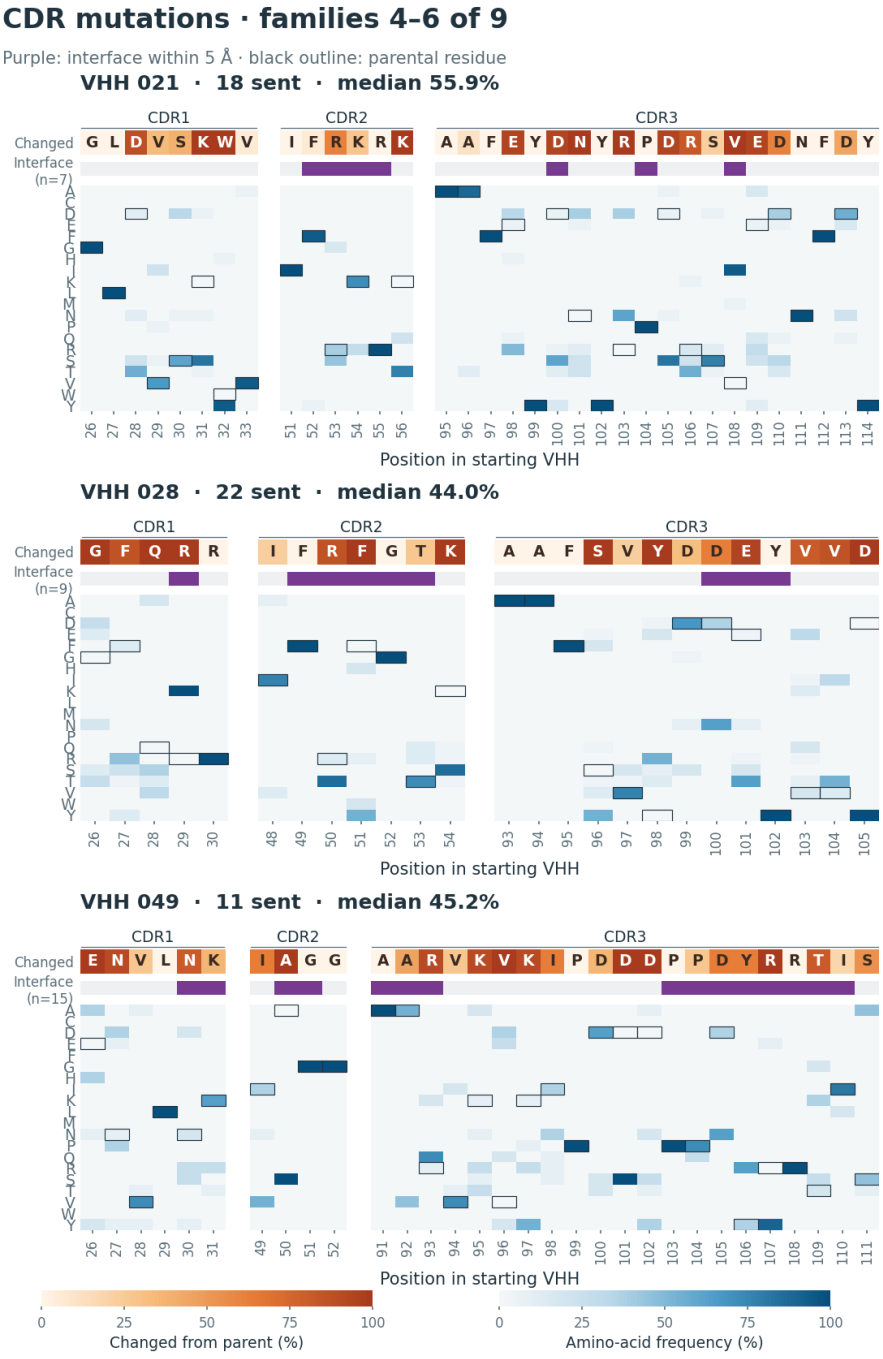


Figure A3B. CDR sequence variation in VHH families 021, 028, and 049, continuing the analysis in Figure A3A. Orange tracks show the percentage changed from the parental residue, blue heatmaps show amino acid frequencies, purple tracks mark interface positions within 5 Å of the target, and black outlines identify parental amino acids.

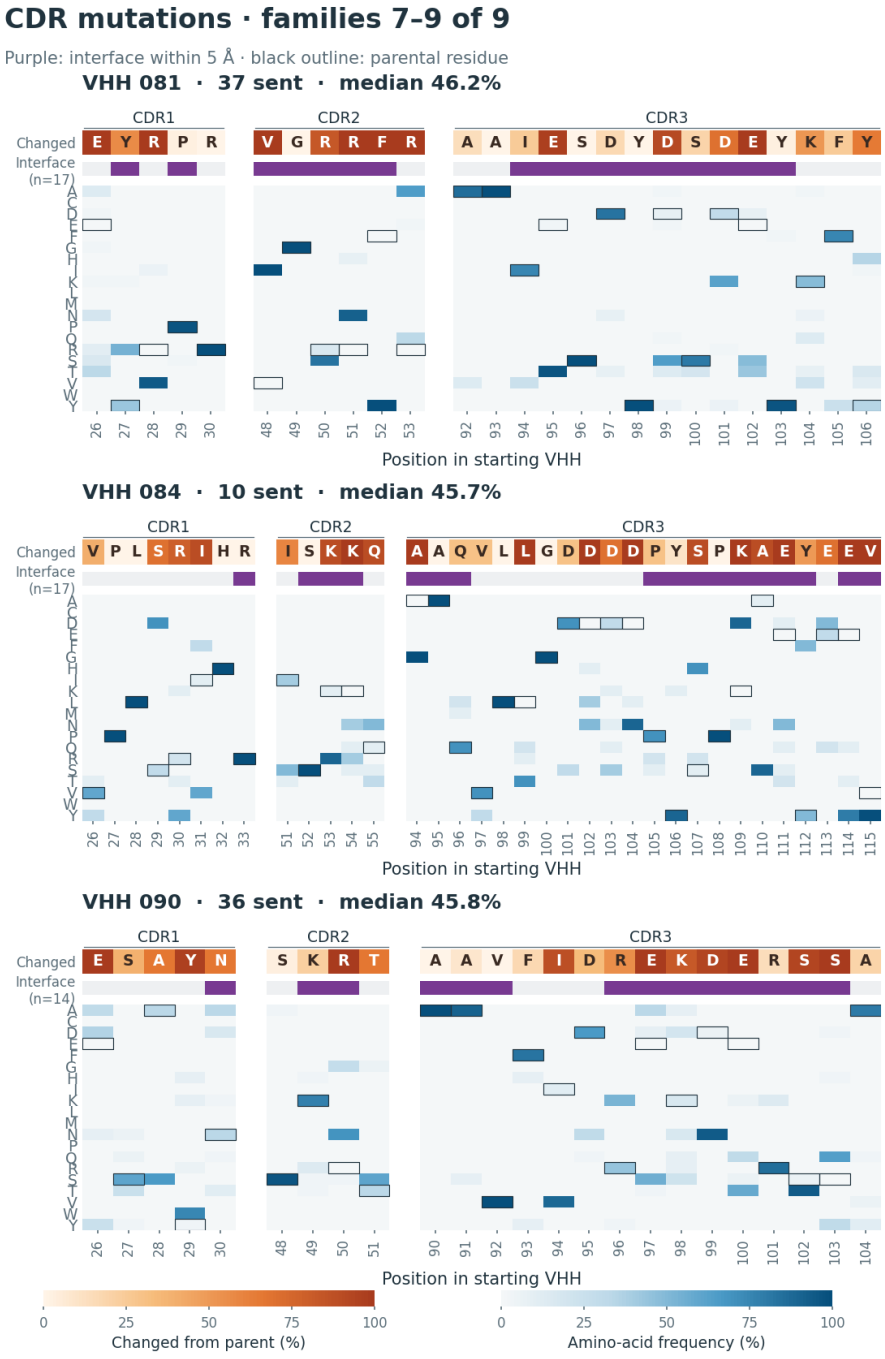


Figure A3C. CDR sequence variation in VHH families 081, 084, and 090, completing the analysis of nine families. Orange tracks show the percentage changed from the parental residue, blue heatmaps show amino acid frequencies, purple tracks mark interface positions within 5 Å of the target, and black outlines identify parental amino acids.

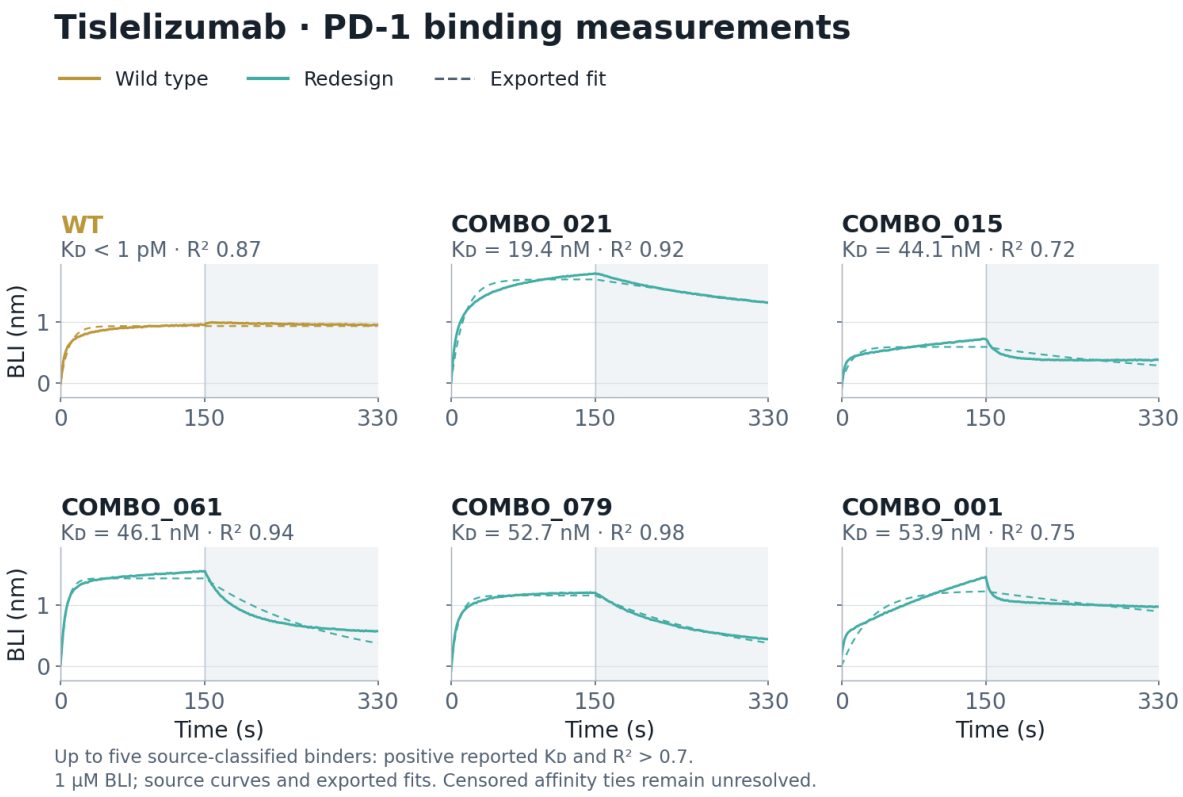


Figure A4A. BLI binding traces for tislelizumab and five selected redesigns against PD-1 in a 1 μM screen. Solid curves show the parental antibody (gold) and redesigns (teal), dashed curves show fitted responses, and labels report KD and R^2; KD values below 1 pM are censored bounds.

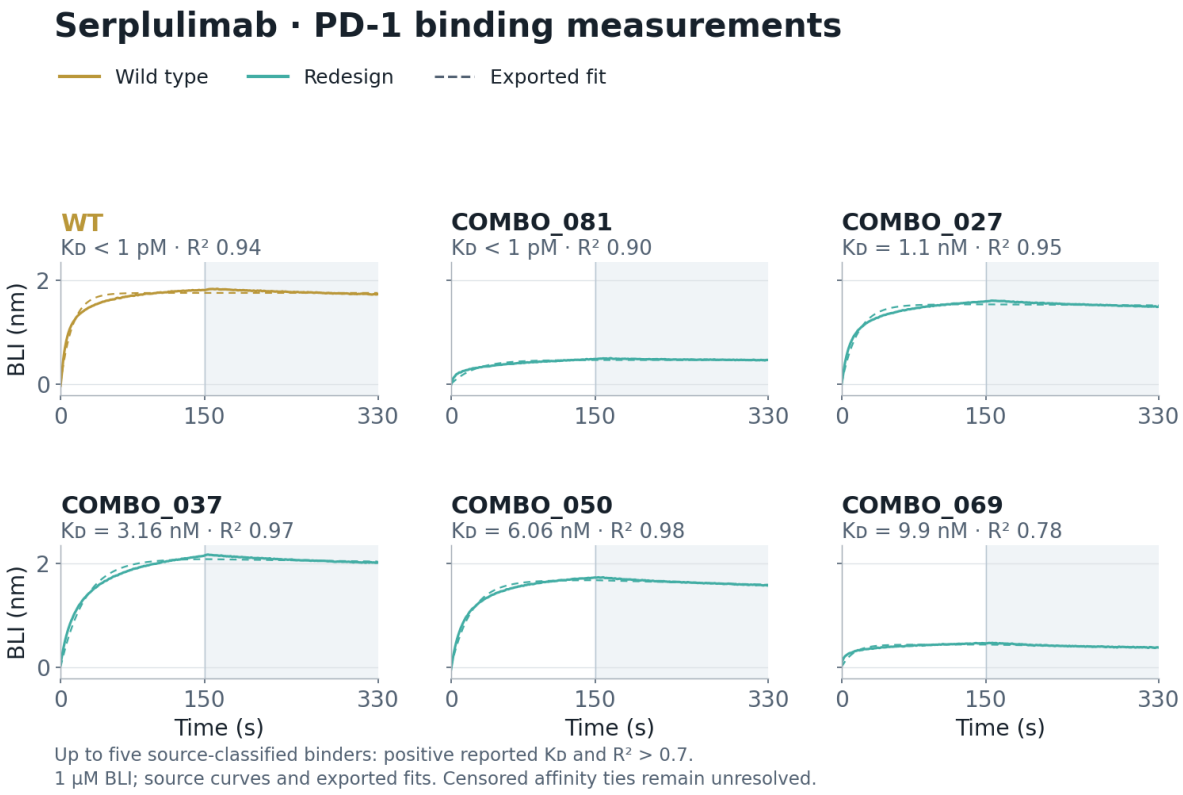


Figure A4B. BLI binding traces for serplulimab and five selected redesigns against PD-1 in a 1 μM screen. Solid curves show the parental antibody (gold) and redesigns (teal), dashed curves show fitted responses, and labels report KD and R^2; KD values below 1 pM are censored bounds.

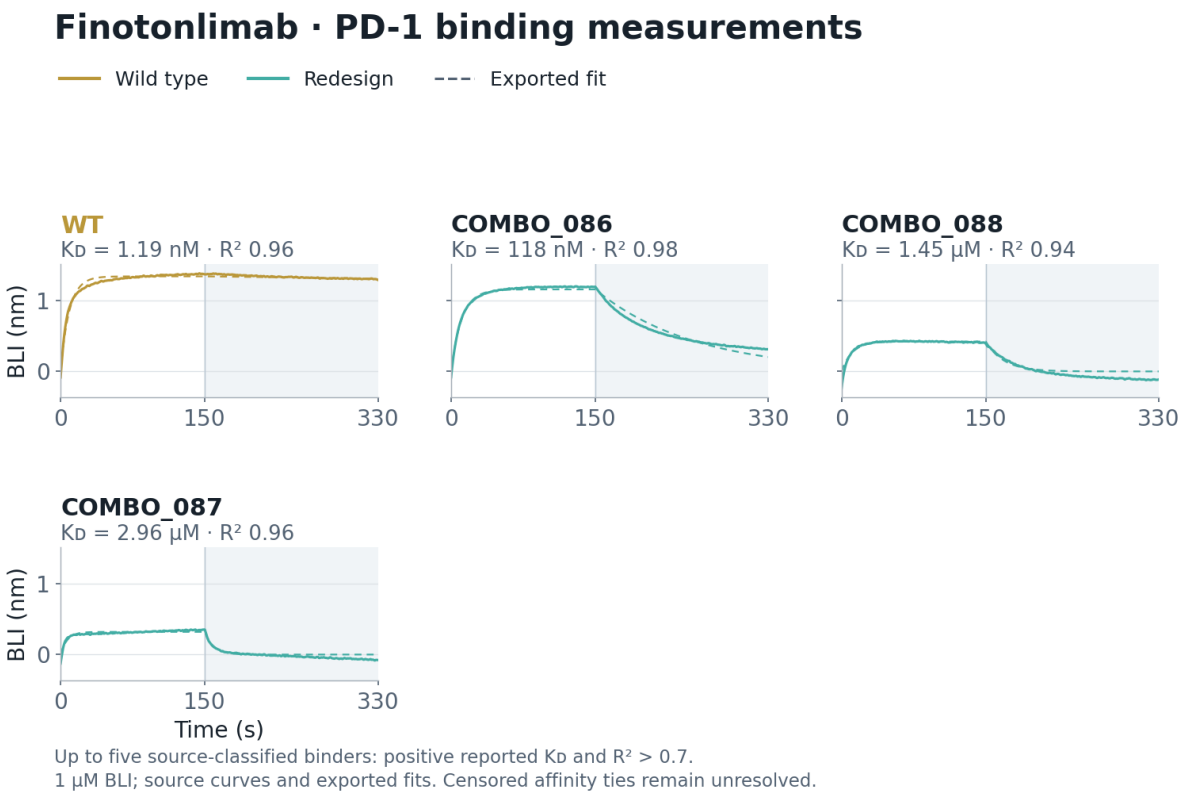


Figure A4C. BLI binding traces for finotonlimab and three selected redesigns against PD-1 in a 1 μM screen. Solid curves show the parental antibody (gold) and redesigns (teal), dashed curves show fitted responses, and labels report the corresponding KD and R^2 values.

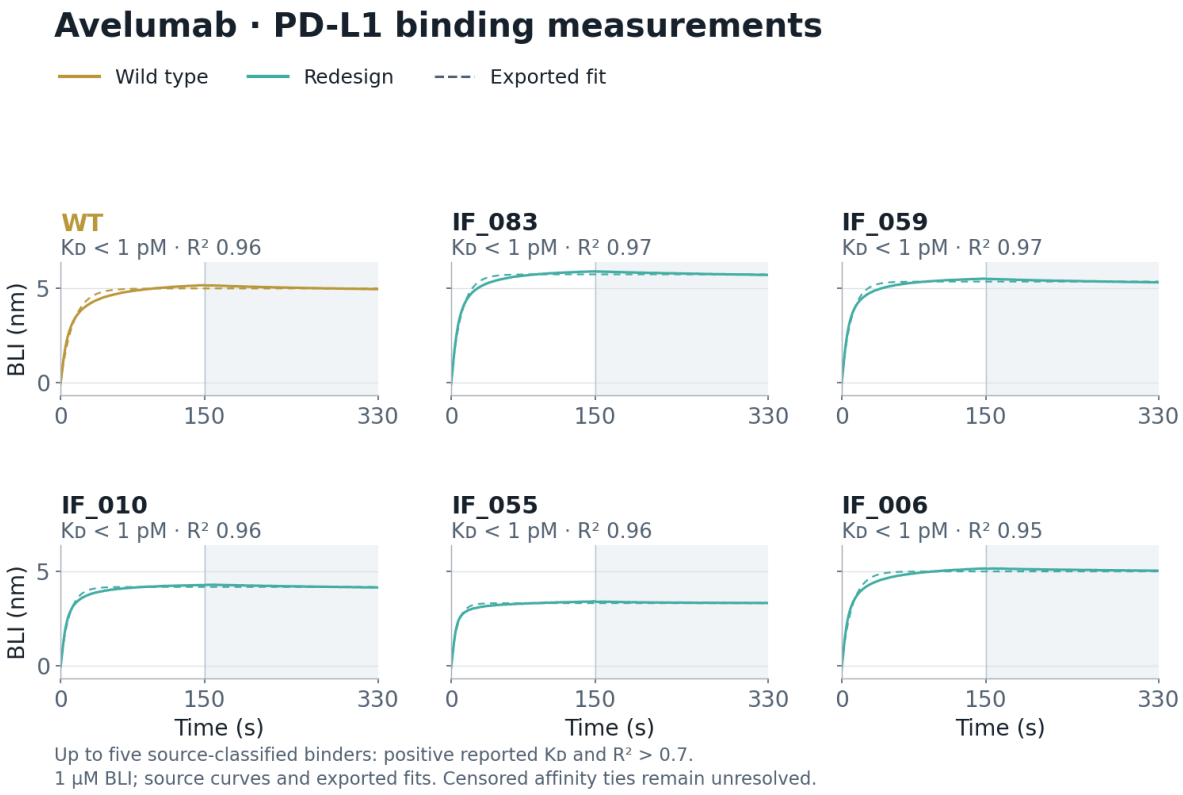


Figure A4D. BLI binding traces for avelumab and five selected redesigns against PD-L1 in a 1 μM screen. Solid curves show the parental antibody (gold) and redesigns (teal), dashed curves show fitted responses, and labels report KD and R^2; all displayed KD values are censored below 1 pM.

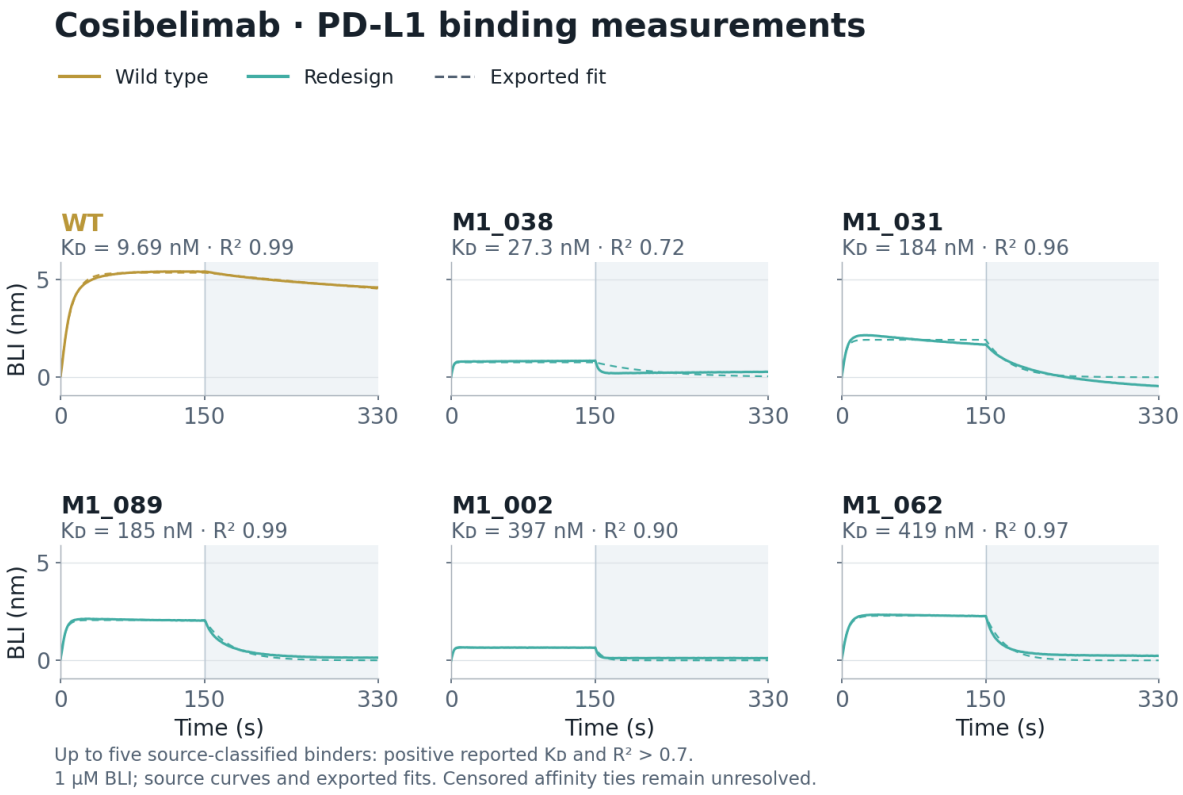


Figure A4E. BLI binding traces for cosibelimab and five selected redesigns against PD-L1 in a 1 μM screen. Solid curves show the parental antibody (gold) and redesigns (teal), dashed curves show fitted responses, and labels report the corresponding KD and R^2 values.

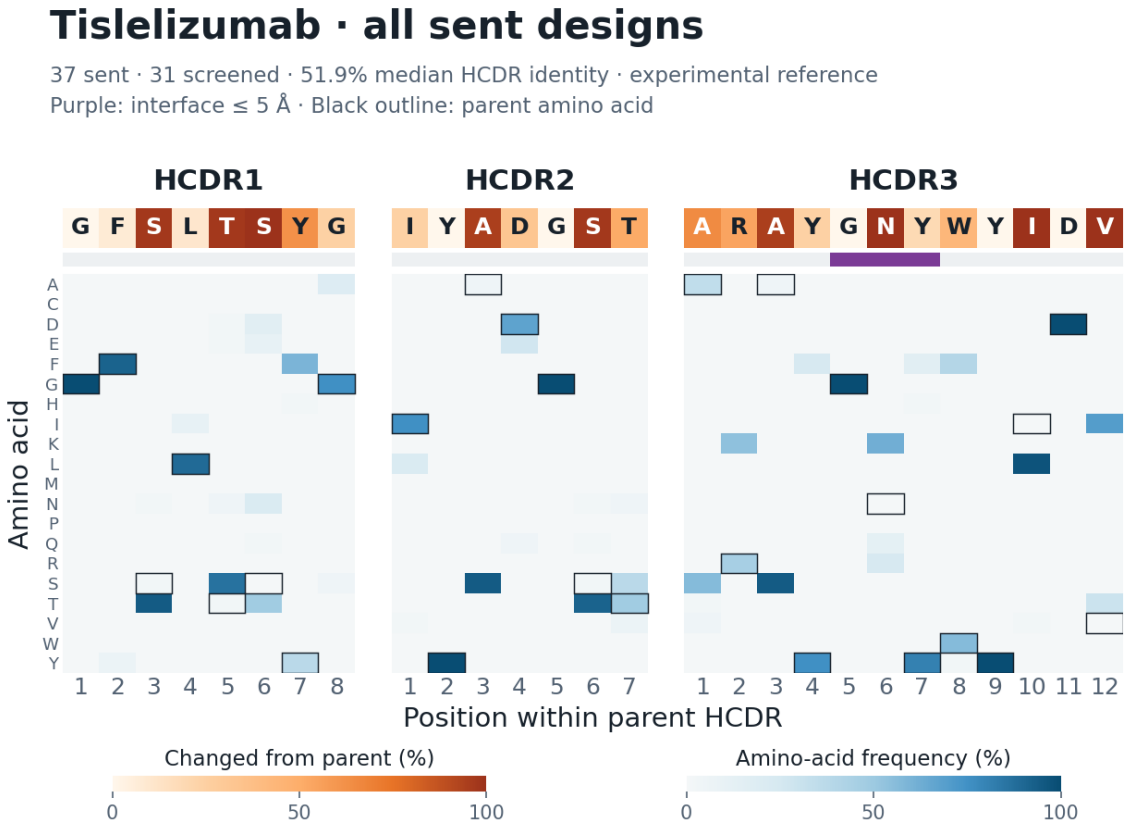


Figure A5A. HCDR sequence variation across all 37 tislelizumab designs sent for experimental testing (31 screened; median HCDR identity to the parent, 51.9%). Blue heatmaps show amino acid frequencies, orange strips show mutation frequencies relative to the parent, black outlines mark parental residues, and purple marks interface positions within 5 Å in the experimental reference complex.

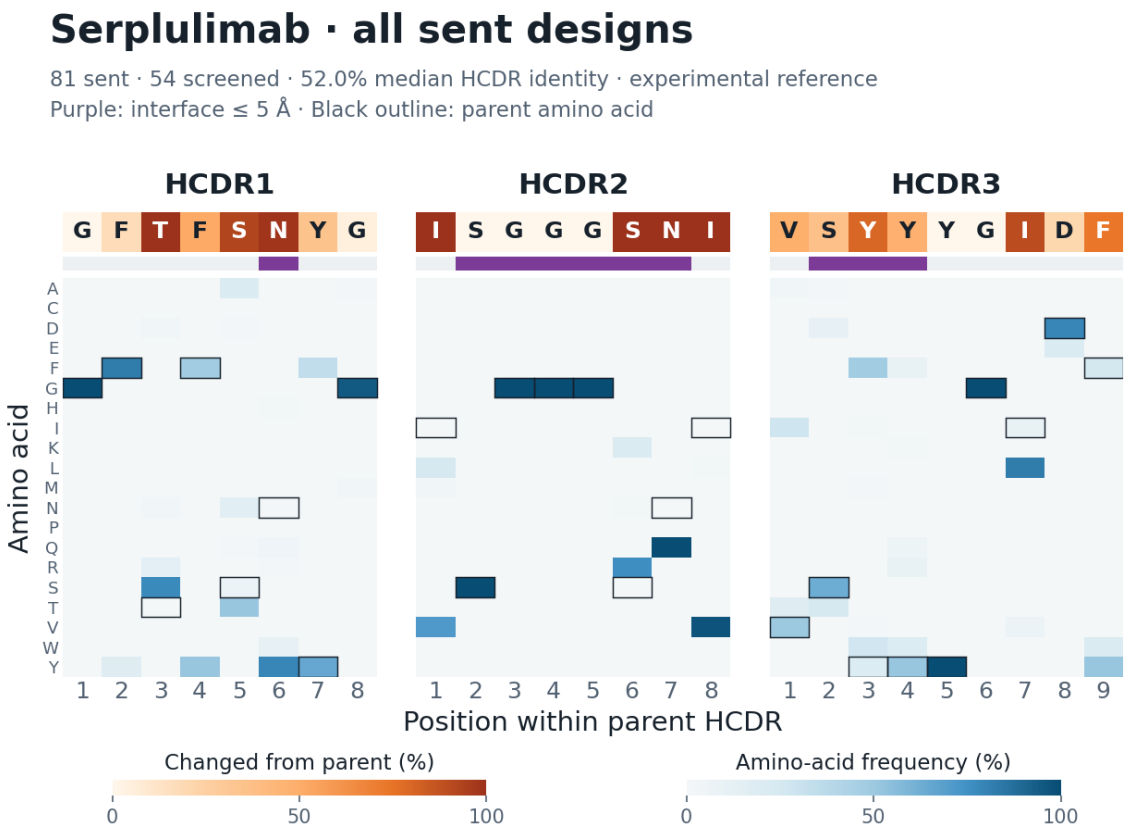


Figure A5B. HCDR sequence variation across all 81 serplulimab designs sent for experimental testing (54 screened; median HCDR identity to the parent, 52.0%). Blue heatmaps show amino acid frequencies, orange strips show mutation frequencies relative to the parent, black outlines mark parental residues, and purple marks interface positions within 5 Å in the experimental reference complex.

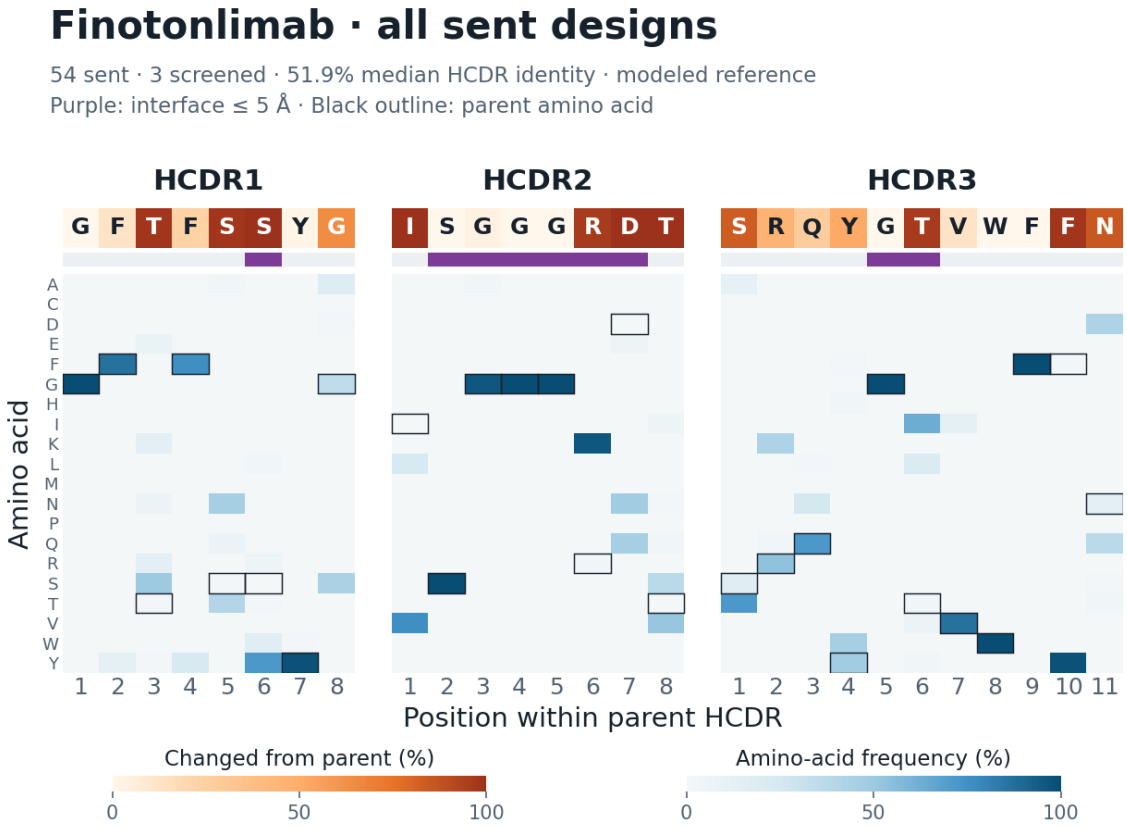


Figure A5C. HCDR sequence variation across all 54 finotonlimab designs sent for experimental testing (3 screened; median HCDR identity to the parent, 51.9%). Blue heatmaps show amino acid frequencies, orange strips show mutation frequencies relative to the parent, black outlines mark parental residues, and purple marks interface positions within 5 Å in the modeled reference complex.

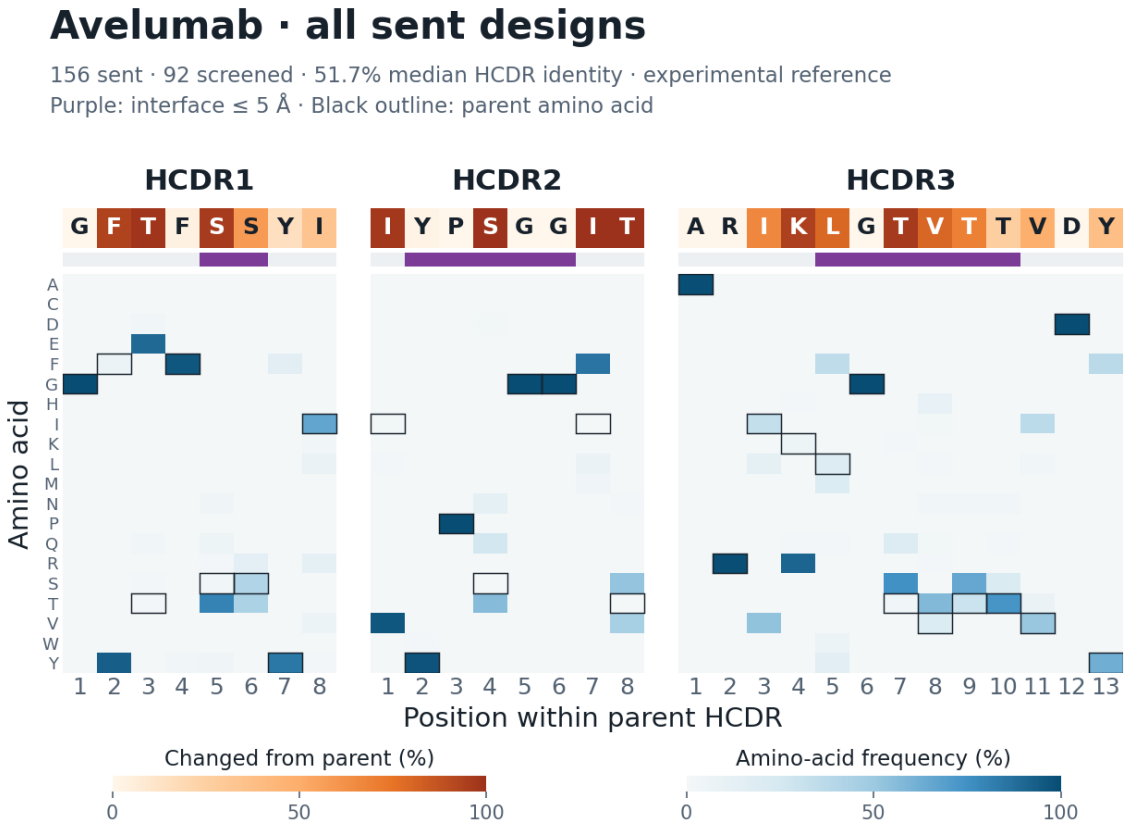


Figure A5D. HCDR sequence variation across all 156 avelumab designs sent for experimental testing (92 screened; median HCDR identity to the parent, 51.7%). Blue heatmaps show amino acid frequencies, orange strips show mutation frequencies relative to the parent, black outlines mark parental residues, and purple marks interface positions within 5 Å in the experimental reference complex.

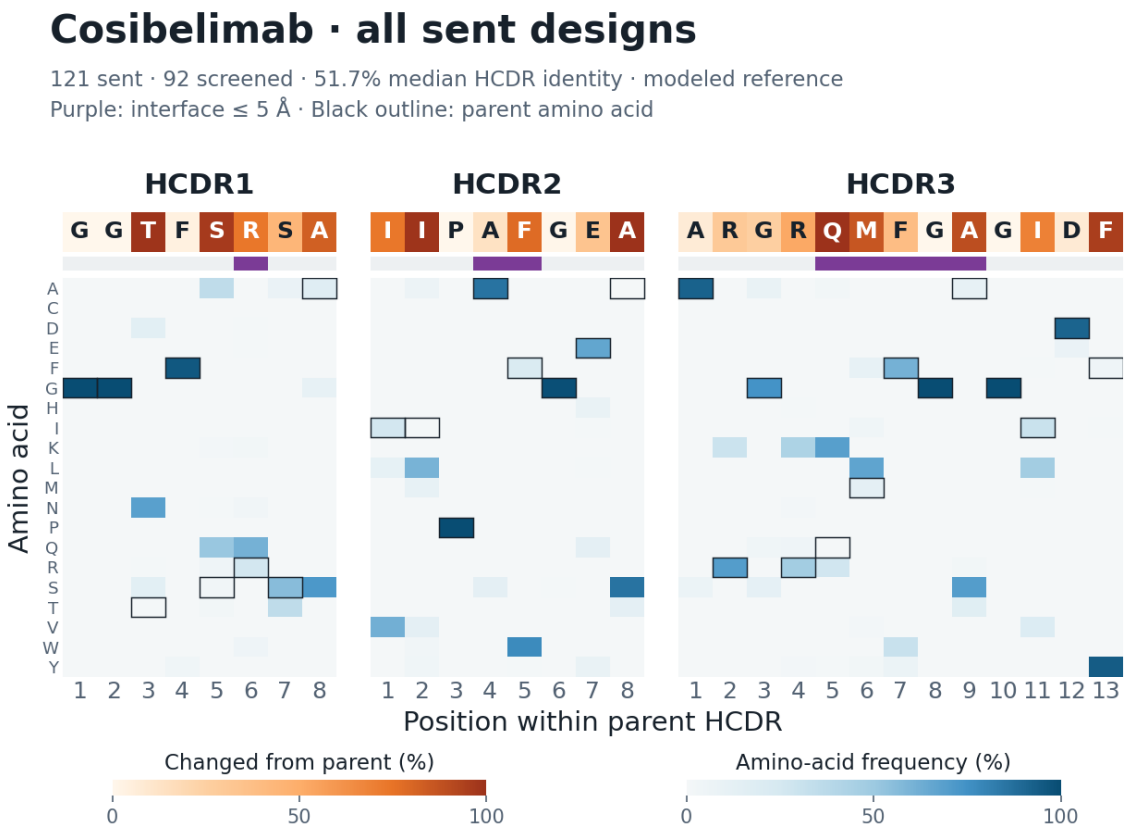
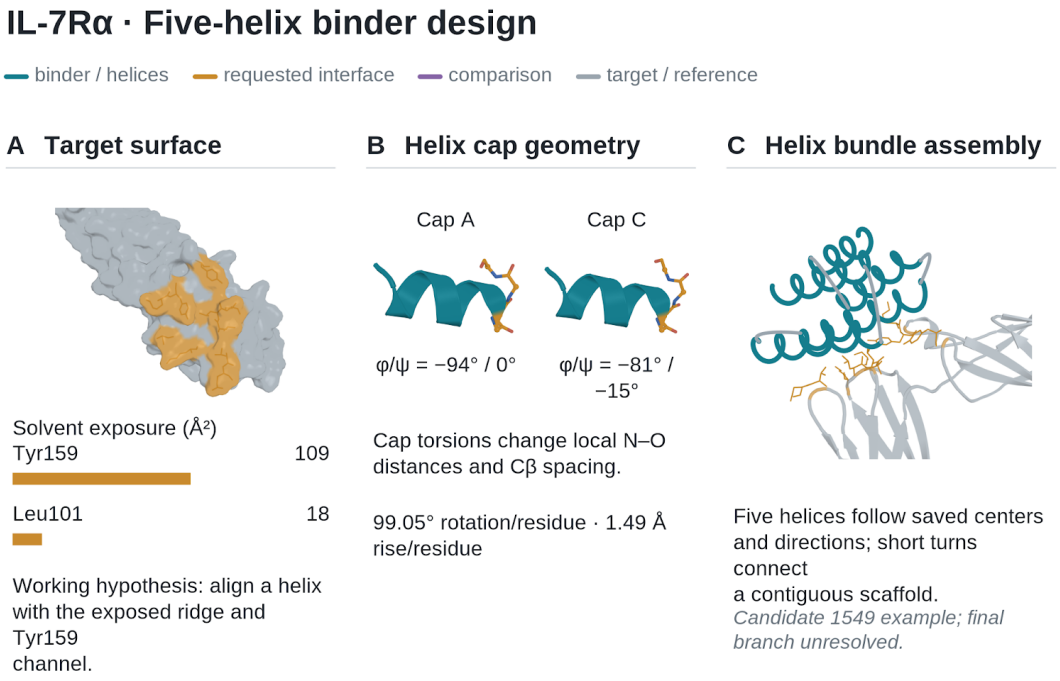


Figure A5E. HCDR sequence variation across all 121 cosibelimab designs sent for experimental testing (92 screened; median HCDR identity to the parent, 51.7%). Blue heatmaps show amino acid frequencies, orange strips show mutation frequencies relative to the parent, black outlines mark parental residues, and purple marks interface positions within 5 Å in the modeled reference complex.

E Further Examples of How Models Solve Protein Biology

E.1 Structure Design

E.1.1 Minibinder Design



VEGF-A · Geometry and backbone selection

binder · target · requested patch · comparison

A Fixed target patch



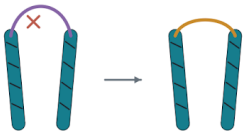
7 requested patch residues

A292 · A249 · A250
A291 · A247 · A246
A252

Preserve both target chains exactly while building one new binder chain.

B Turn-test revision

Initial test	Revised test
acceptor offset −5	acceptor offset −4
φ/ψ (−66°, −38°)	φ/ψ (−61°, −30°)
0 / 0 / 0 accepted	**36 / 54 / 139 accepted**

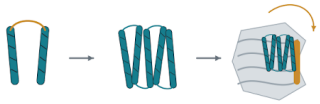


Spacer lengths 2 / 3 / 4

Both settings changed together; neither effect is isolated.

C Backbone growth and docking

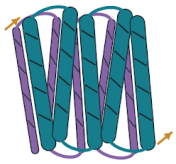
Helix-pair fragments	Compact bundles	Dock to the patch
splice connectors	screen packing	rotate + translate



Checkpointed docking results guided flexible refinement.

D Refinement and selection

Flexible refinement



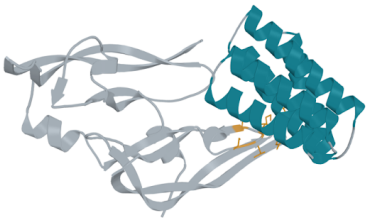
Torsion + pose gradients

Candidate scores

res20	119.817
res4	119.696
res11	**119.468**

Packing, hotspot checks, and geometric inspection informed the res11 choice.

E Final complex and checks



Model-time checks

107-residue Ala/Gly backbone

Target coordinates preserved

Closest target–binder: 3.028 Å

Mean peptide C–N: 1.329 Å

All 7 hotspots within 5.378 Å

Evaluation

iPSAE	**0.658**
Target-aligned Cα RMSD	**1.352 Å**

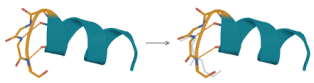
VEGF-A · Sequence design

binder / scaffold · target interface · alternative · target / reference

A Helix cap refinement

Before local refinement

After refinement · gray = before



refine

Local cap N–O distance: 2.29 → 2.84 Å.

B Scaffold placement and screening

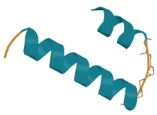
Sample placements



Score contact and overlap

target contact · steric overlap · internal packing

Refine promising backbones after placement screening.



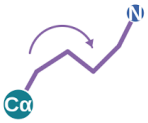
Late candidate fragment · 35 residues

Local frame; placement unresolved.

C Residue and sidechain search

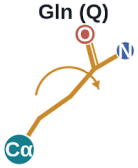
Choose a residue, then test sidechain angles

Lys (K)



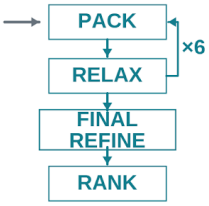
long charged sidechain

Gln (Q)



polar amide sidechain

rotamer choices

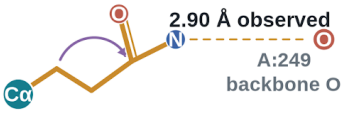


Rank completed designs

100 jobs completed on 11 workers · controlled score perturbations · final refinement

D Gln62 rejection and Lys62 confirmation

A close contact was not enough

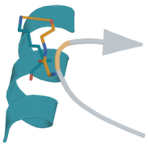


Gln62 second rotamer angle

−112.4° vs −80.8° reference

REJECT Q62

The rollout returned to a Lys62 candidate.



Final Lys62 NZ–A:250 O: 3.01 Å

A:249 O remained flagged by the custom polar diagnostic.

E Final complex and evaluation

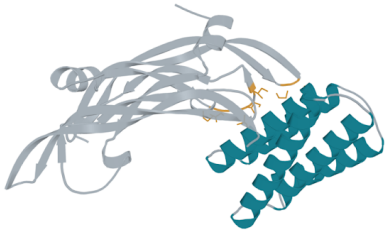
120-residue binder

Target preserved

C–N: 1.328–1.330 Å

Minimum interface distance: 2.840 Å

The final check noted a potentially strained Trp65 rotamer.



iPSAE min **0.595** Target-aligned binder Cα RMSD **1.650 Å**

TrkA · Side-chain annealing and refinement

binder · target · requested patch · comparison

A Fixed target patch



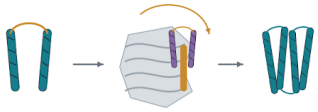
5 requested patch residues

X297 · X333 · X343
X296 · X302

Keep TrkA unchanged while placing one compact binder against the requested patch.

B Hairpin placement and growth

Helix-pair fragments | **Focused placement** | **Packed core**

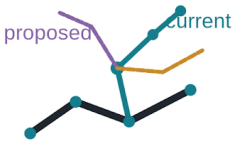


sample connectors | rotate + translate | favor compact cores

Docking test after repair: 596 fitted hits; 173 passed its custom anchor-bonus check.

C Identity and rotamer annealing

One rotamer change



Energy update

ΔE = option change
+ incident pair changes
+ composition change

uphill: exp(−ΔE/T), within guards

T = 0.89 — T = 0.008

12 configured temperatures

Three design/relax cycles ran. The local gradient check reported a maximum error of 1.40×10^{-8}

D Candidate selection and cleanup

48 completed redesigns

Candidate 24	Candidate 34
−65.916	**−55.962**
lower custom score	selected: helix/core preference

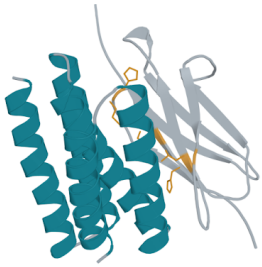
Largest custom checked overlap



0.337 Å → **0.101 Å**

RMS movement 0.064 Å; stopped at the 390-iteration limit.

E Final complex and checks



Model-time checks
118-residue binder
Target records preserved
Peptide C–N: 1.331–1.334 Å
53 Cα pairs within 8 Å
Total SASA loss: about 2392 Å²

Evaluation

iPSAE	**0.845**
Target-aligned Cα RMSD	**0.893 Å**

IL-17A · Six-helix binder design

target · binder · connector / alternative · turn / strand · interface

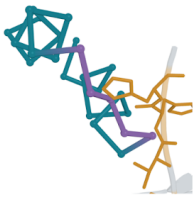
A Strand placement



Five Cα positions from a strand-pairing test against the IL-17A tail.

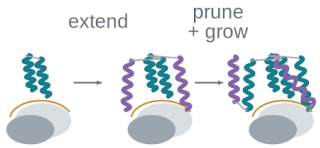
B Short-strand connector

Separate seed



A 25-Cα seed. Purple marks the five-position connector, including two
E-coded strand-like positions.

C Six-helix growth and pruning



C/D seed **Guided extensions** **ABCDEF family**

B-growth: 3,729 candidate outputs / 3,740 jobs

Extend helices toward guide endpoints, then prune redundant candidates.

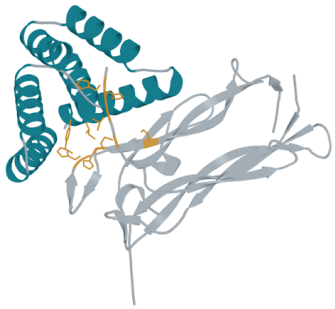
D Connector polar contacts

Final structure shown for reference



The Thr64 restraint targeted 2.89 Å; the final distance was 3.16 Å.

E Final complex



Minimum target-group iPSAE

0.522

Target-aligned binder Cα RMSD

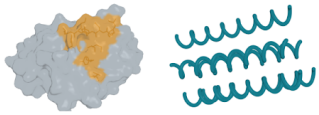
1.975 Å

108-residue binder • short connector and six-helix clamp

PD-L1 · Loop assembly and interface refinement

binder / helix · target / reference · matched / interface · comparison

A Patch and scaffold frame



Target patch · Five nominal helix guides

Target-relative frame; five straight helix templates precede trimming and loop closure.

Nominal cross-section



B Loop position and orientation



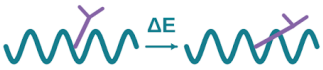
Amber: matched · Gray: mismatched alternatives

3D displacement · 3 × relative rotation · loop preference

Library lookup + trimming cost choose each connector

Match both endpoint position and chain direction.

C Sidechain annealing



current rotamer · proposed rotamer

ΔE ≤ 0: accept

ΔE > 0: accept with exp(−ΔE/T)

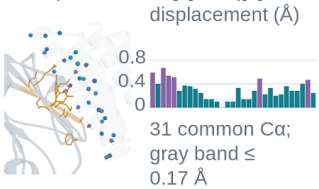


Keep the best state, then run two greedy cleanup sweeps.

Example search schedule

D Late interface refinement

0.1 Å precision · ed[8] → fq[3] displacement (Å)



31 common Cα; gray band ≤ 0.17 Å

Colored: fq[3] Cα; gray: ed[8]. Faint cartoon: exact final scaffold for orientation.

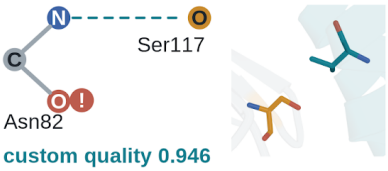
Selected late path

ed[8]
↓ 0.19 Å RMS
fin[3]
↓ 0.34 Å RMS
fq[3]
↓ 0.10 Å RMS
final

Final 0.10 Å shift is near rounding uncertainty.

E Polar tradeoff

Before two-site repack · After repack + refinement



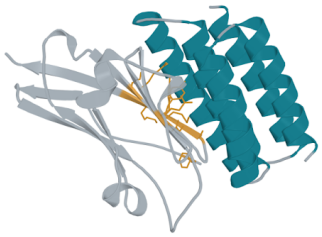
custom quality 0.946

Asn82 O and Asp103 O flagged · Val82 · Ser117

Val103 also selected; Ser117 Oflagged

Removing two binder polar groups also removed modeled Asn82 donation. Binder warnings changed while a target polar warning appeared. Two residue choices and coordinates changed together.

F Final complex and evaluation



112 residues · five helices · target preserved

Continuous peptide geometry; closest target–binder heavy-atom distance 2.73 Å.

iPSAE	**0.898**
Target-aligned binder Cα RMSD	**2.489 Å**

PD-1 · Strand pairing and binder design

binder / guide · β strand · target interface · comparison · target / reference

A Local helix geometry



Eight printed Cα positions
1.49 Å rise per residue
98.90° rotation per residue

B Antiparallel strand anchoring



Final structure shown for reference

PD-1 strand direction



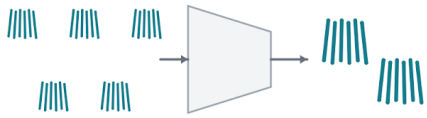
Binder strand direction

Opposite directions define antiparallel pairing.
Seeding: loops of 4–8 residues, extended bins and bulges,
three anchors, then relaxation.

C Bundle growth and candidate guides

Expand candidate bundles

Retain a smaller beam



25,500 generated **1,500 retained**

Observed growth step 4 → 5

Two inspected redesign alternatives

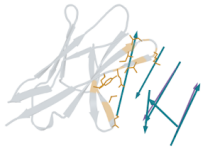


candidate 0 candidate 263

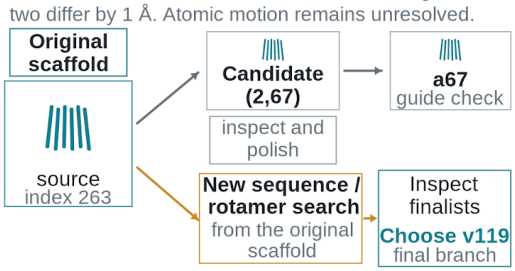
Guide means rounded to 1 Å

Candidate 263 is a redesign; later branches start from its original scaffold.

D Guide polishing and branch selection

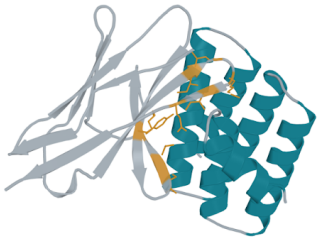


Purple: (2,67) · Teal: a67
Guide means rounded to 1 Å: 10 of 12 unchanged; two differ by 1 Å. Atomic motion remains unresolved.



a67 and v119 are separate redesign branches.

E Final complex and evaluation



111-residue binder

Target preserved
545 atom pairs below 5 Å
C–N: 1.330–1.332 Å

Final reciprocal Z15:Y134 N–O distances: 2.88 and 2.91 Å

iPSAE min	**0.730**
Target-aligned binder Cα RMSD	**1.884 Å**

PD-1 · Sheet assembly

hairpin / helix · strand / bridge · alternate hairpin · PD-1 / before

A β-Hairpin options

LG · 16 residues



eD · 16 residues



ADG · 17 residues

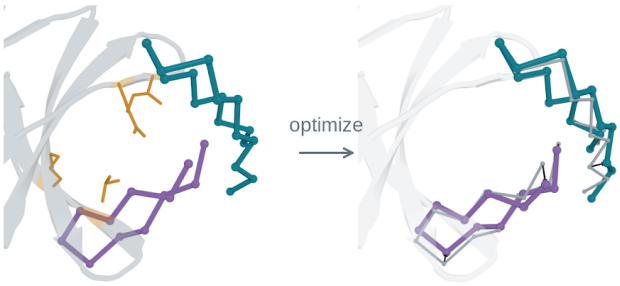


Dashed ellipses mark the varied turn.

B Hairpin combination and relaxation

Before combination

After optimization · motion overlay



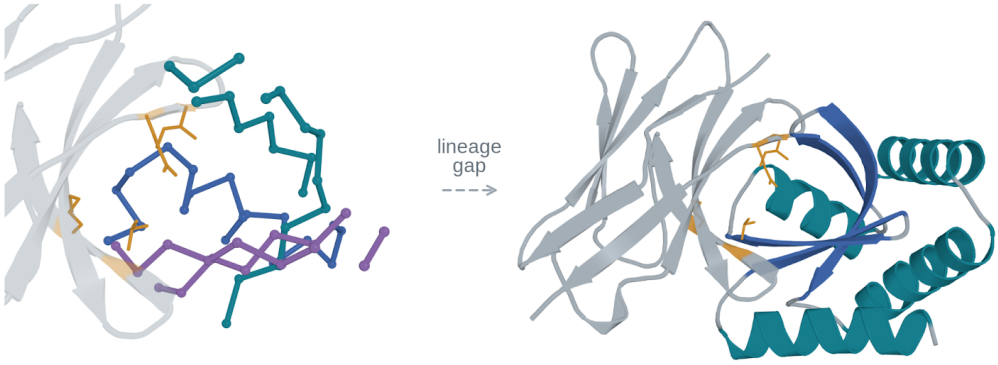
29 Cα positions · RMS shift 0.9 Å · Largest shift 2.1 Å

Coordinates rounded to 0.1 Å; gray = before, color = after.

C Bridge candidate and final binder

Bridge candidate · 55 Cα

Final binder · 115 residues



4 strand runs + 4 helix runs · Seven target-pairing N–O distances: 2.86–3.15 Å

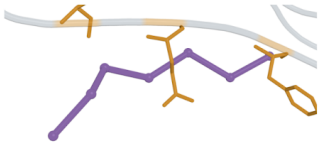
iPSAE 0.664 · Target-aligned binder Cα RMSD 1.666 Å

CD40 · Five-helix binder design

target · binder · connector / alternative · turn / strand · interface

A Strand placement

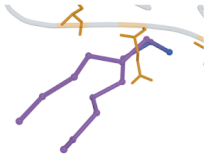
Exploratory test



Seven Cα positions after revising the target-pairing restraints.

B Hairpin alternative

Separate hairpin example



A separate 14-Cα hairpin optimization. Blue marks a positive-φ turn position.

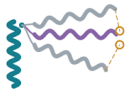
C W-shaped scaffold sampling

Configured W sites



Five configured cross-section sites

Extension choices

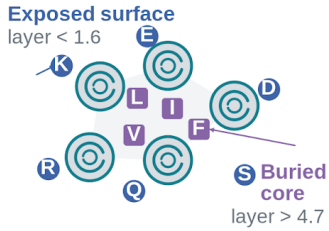


Endpoint fit + Gumbel noise

choose min(error² − 21 × Gumbel)

The recorded search combined endpoint fit with Gumbel noise. The test returned five bundles, but the rollout questioned their shared cgpn turns.

D Directional core packing

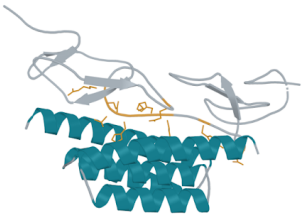


Letters are examples from the recorded allowed sets.

Packing ran after the model corrected an array index.

Away from the target (tlay < 0.4), low burial favored polar/charged options and high burial admitted hydrophobic options. This is a design prior.

E Final complex



Minimum target-group iPSAE

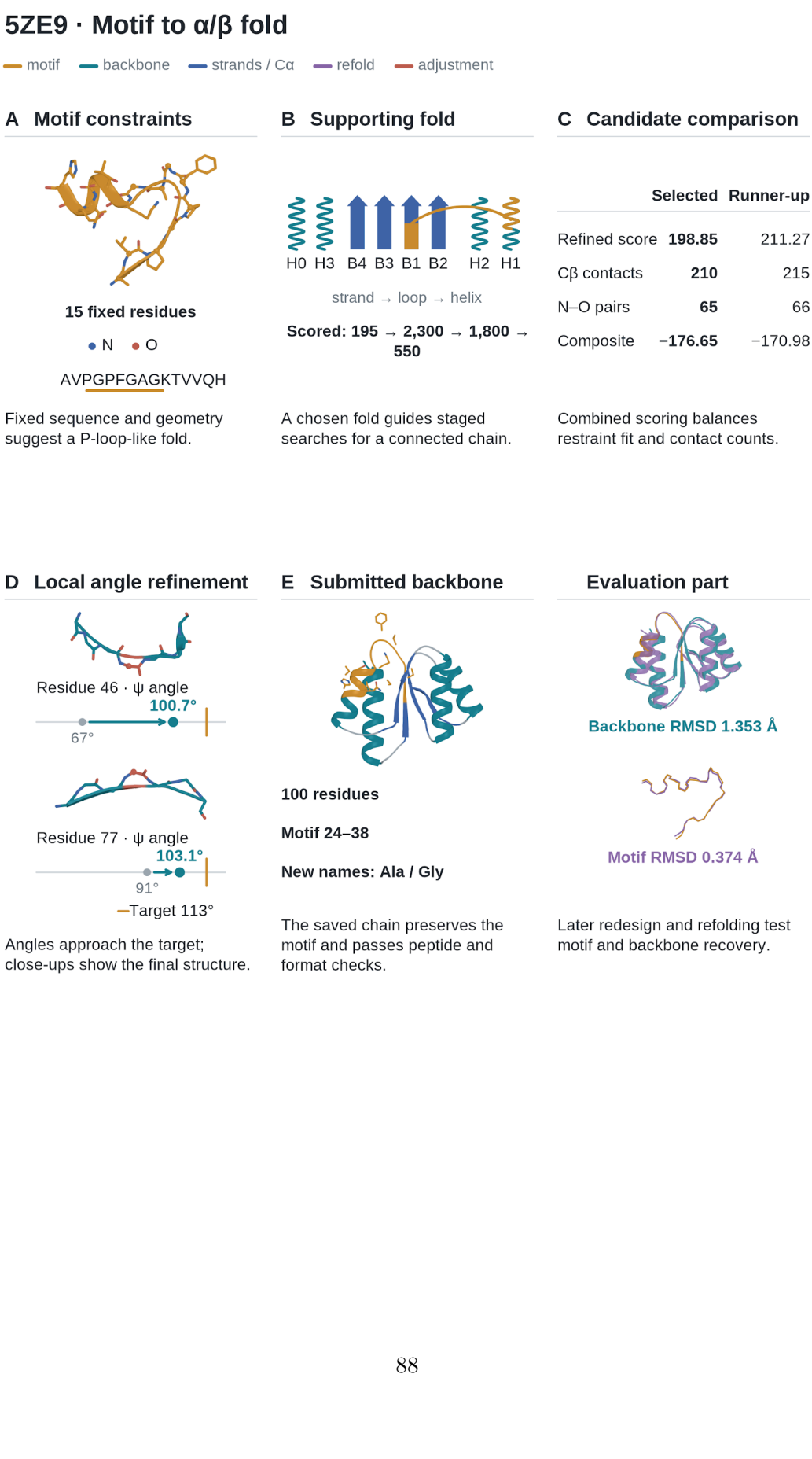
0.760

Target-aligned binder Cα RMSD

1.854 Å

120-residue five-helix binder • refined from candidate 113

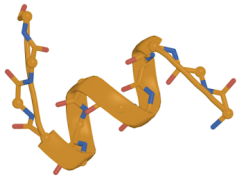
E.1.2 Motif Scaffolding



6E6R · Helix to protein bundle

motif · backbone · strands / Cα · refold · adjustment

A Motif constraints



Backbone	**13 residues**
Fixed identities	**none**
Positive φ turns	**+84° / +76°**

N O

Geometry suggests a capped helix; all motif residue identities remain designable.

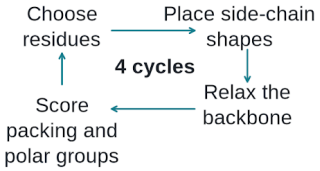
B Layout search

800 → **218** → **20**

layout draws · accepted specifications · constructed examples

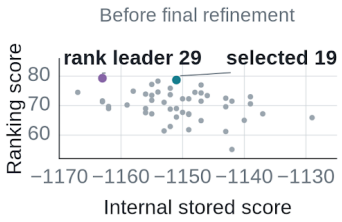
Target geometry guides this exploratory batch; its final ancestry remains uncertain.

C Packing cycles



Temporary side chains test packing while repeated refinement adjusts the backbone.

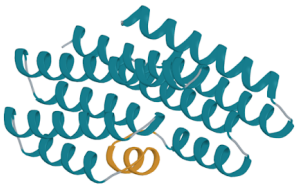
D Finalist comparison



After refinement: exposure 19 = 2.6; 29 = 3.1

Variant 19 is selected; variant 29 scores better on contact metrics.

E Submitted backbone



200 residues

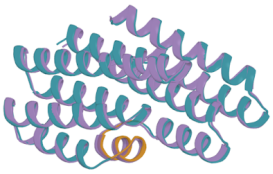
800 backbone atoms

C–N links: 1.330–1.334 Å

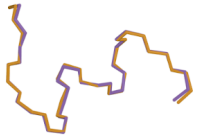
Motif 104–116

The saved backbone preserves the motif and passes chain and geometry checks.

Evaluation part



Backbone RMSD 0.662 Å



Motif RMSD 0.183 Å

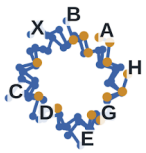
Later redesign and refolding test motif and backbone recovery.

2RKX · Eight fragments to a barrel

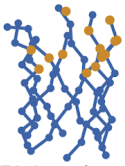
motif · backbone · strands / Cα · refold · adjustment

A Early barrel core

Top view



Side view

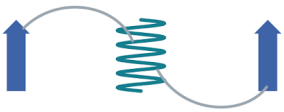


70 Cα points

Supplied fragments guide a recorded Cα prototype, not the final barrel core.

B Connector search

Loop → helix → loop



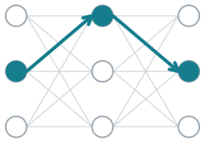
anchor 1 outer helix anchor 2

8 libraries · revised G/F sector

Budget: 11,000 per module

Loop–helix–loop searches connect strand ends; the irregular G/F sector receives further revision.

C Compatibility and length



Previous Current Next

Pair-repulsion proxy

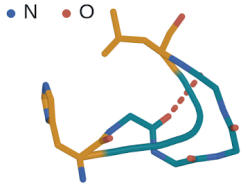
2,598.9

~0 after joint choice

Intermediate linker: 189 residues

An intermediate connector set reduces one clash penalty; a different combination supplies the final length.

D Local contact refinement



O35···O38

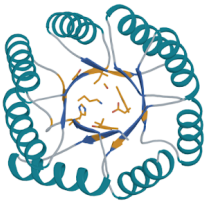
First adjustment
target 3.042 Å
measured 2.982 Å

↓

Retuned adjustment
target 3.016 Å
final 2.941 Å

Final markers locate the retuned contact; fixed motif strain remains.

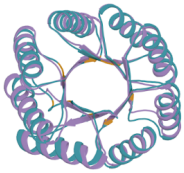
Evaluation part



225 residues

8 motif pieces preserved

C–N links: 1.323–1.340 Å



Backbone RMSD 1.439 Å



Motif RMSD 1.180 Å

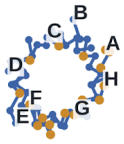
Later redesign and refolding test motif and backbone recovery.

3B5V · Build a 200-residue scaffold

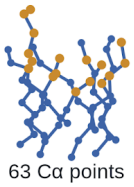
motif · backbone · strands / Cα · refold · adjustment

A Early strand core

Top view



Side view

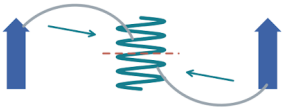


63 Cα points

Supplied fragments guide a recorded Cα core before cropping and connector construction.

B Connector assembly

Join the two halves



anchor 1 outer helix anchor 2

Helix lengths: 9–16 residues

Connector halves grow from neighboring anchors and meet near helix midpoints.

C Exact-length assembly

50	+ 150	= 200
core residues	connector / lead residues	total

9,369	1,480	1
exact-length link sets	completed assessments	selected layout

Compatibility search enforces exactly 200 residues while comparing connector combinations.

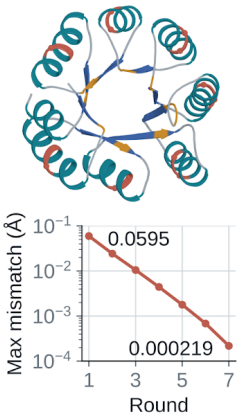
D Packing candidate comparison

	Chosen variant 7	Variant 4
Internal score	**−817.2**	−829.3
Unpaired groups	**18**	24
Weighted estimate	**8.5**	12.0

Full rationale unavailable

Variant 7 advances after comparing internal scores and estimated unpaired polar groups.

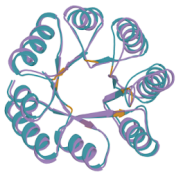
E Seam closure



Fixed C140–O140: 1.177 Å

Computed copies converge; markers annotate seams on the final structure, not archived gaps.

Evaluation part



Backbone RMSD 1.239 Å



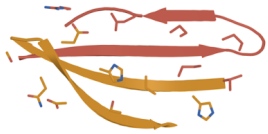
Motif RMSD 1.033 Å

Later redesign and refolding test motif and backbone recovery.

5IUS · Two fragments to a β sandwich

motif · backbone · strands / Cα · refold · adjustment

A Motif constraints



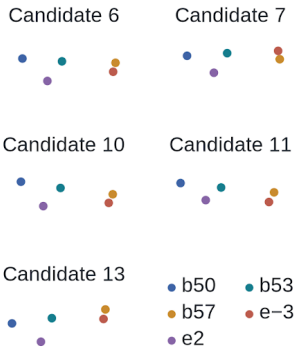
- **A · 20 residues**
  XHXVXHXEXXSG
  QXDTXXAX
  10 fixed identities
- **B · 22 residues**
  XXXXXGXIXLXX
  KIXIXEXXRX
  8 fixed identities

X = redesignable identity

N · O · S

Two fixed geometric fragments must be joined within one compact chain.

B Sheet placement comparison



Local landmark key · no connections

Sparse Cα landmarks compare alternative sheet placements, not complete structures.

C Loop candidate comparison

	EBABP carried forward	DPGBB alternative
Internal score	**−160.49**	−163.14
Tyr geometry factor	**0.886**	0.763

Loop matching retries without stale restraints; the Tyr ranking atom is temporary.

D Joined Cα prototype



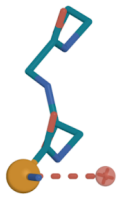
100 Cα points

Motif segments

EF loop 72–76

Recorded Cα points show an exploratory joined prototype before final packing and refinement.

E Terminal refinement



- Final residue-1 Cα
- Configured target

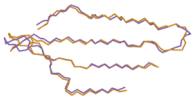
2.841 Å away

Terminal refinement pulls the chain toward a configured guide point.

Evaluation part



Backbone RMSD 1.178 Å



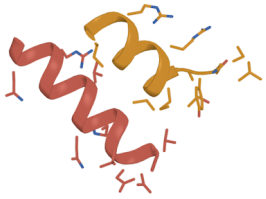
Motif RMSD 0.866 Å

Later redesign and refolding test motif and backbone recovery.

7DGW · Two helices to a bundle

motif · backbone · strands / Cα · refold · adjustment

A Motif constraints



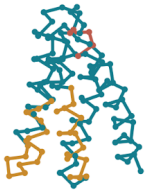
A · 15 residues
NYPLTPVMRAAA
RAI
15 fixed identities

B · 15 residues
DLELARQALAAA
RQA
15 fixed identities

N · O · S

Two fixed helical segments must fit within one compact, continuous chain.

B Early bundle prototype



125 Cα points

Overlap 50–54

Closure RMS
0.152 Å

This exploratory Cα prototype reaches iteration limits and retains closure error.

C Layout and packing search

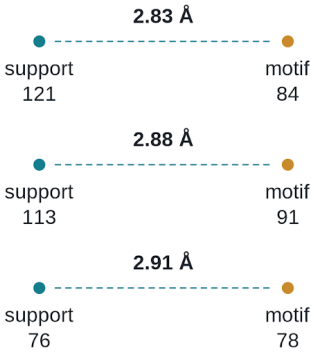
34 → **240** → **8,160**

generations · offspring per generation · completed evaluations

Plus initialization; repeated sequences possible

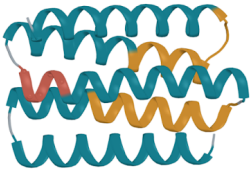
Helix layouts vary while a genetic search mixes side-chain choices and refines backbones.

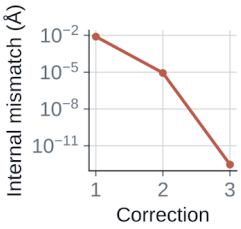
D Polar-group support



Internal support distances use temporary side chains omitted from the saved PDB.

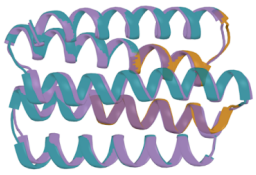
E Backbone overlap correction





Internal mismatch: 0.00796 → 2.92×10^{-13} Å
Duplicated Cα positions converge numerically; markers locate the overlap on the final structure.

Evaluation part



Backbone RMSD 0.696 Å



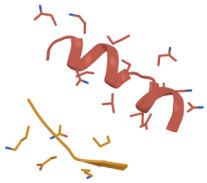
Motif RMSD 0.662 Å

Later redesign and refolding test motif and backbone recovery.

4JHW · Build around a cis peptide

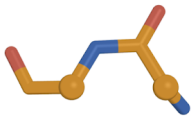
motif · backbone · strands / Cα · refold · adjustment

A Cis-peptide motif



A: 7 residues · B: 17 residues

N O S



Fixed cis Lys–Glu

ω ≈ −2.33°

The supplied motif includes a fixed cis Lys–Glu peptide and mixed structural geometry.

B Geometry-score repair

Malformed restraint

atom1	atom2	atom3	178°

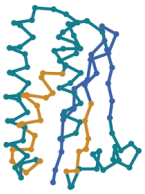
Corrected restraint

atom1	atom2	atom3	atom4	178°

Derivative difference: 1.3×10^{-12}

A missing atom index is repaired; finite differences confirm the corrected gradient.

C Partial scaffold construction



114 Cα points

Partial motif

17–23 / 52–68

Connector searches build partial layouts; this recorded Cα candidate has unresolved final ancestry.

D Disulfide fitting

S–S target 2.03 Å



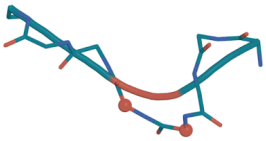
internal Cys site internal Cys site

Internal fit costs

−69 / −69	−69 / 180	180 / 65
12.83	**2.53**	**8.52**

Different starting conformations test a modeled disulfide; sulfur atoms are omitted from export.

E Exit-junction check



B-exit junction 110–111



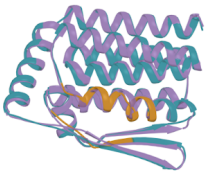
Modeled H–H: 1.63 Å

Nominal threshold: 1.92 Å

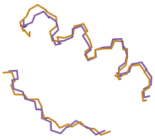
ω ≈160.2°

Final markers locate the junction; its short H–H distance uses temporary hydrogens.

Evaluation part



Backbone RMSD 0.882 Å



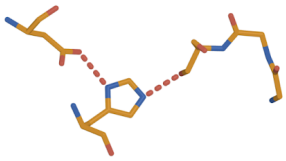
Motif RMSD 0.865 Å

Later redesign uses Ile65/Gly110; the proposed disulfide is absent.

4XOJ · Build around a polar motif

motif · backbone · strands / Cα · refold · adjustment

A Polar motif



Asp–His	**2.752 Å**
His–Ser	**2.684 Å**

N · O · S

Supplied Asp–His–Ser geometry suggests a triad; catalytic function remains untested.

B Bridge optimization

Before



Closure penalty 7,329.5

After



Closure penalty 8.1

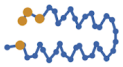
An exploratory bridge closes between fixed anchors; peptide strain remains.

C Helical route comparison

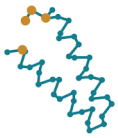
37 residues



39 residues



43 residues



Rounded Cα · shared projection · individual scale

Exploratory Cα traces test three helical routes around the motif.

D Bundle candidate comparison

	Chosen 18 Candidate 1	
Backbone	**−595.04**	−617.48
Sequence score	**−37.94**	−25.65
Modeled H–O	**2.373 Å**	3.078 Å
Composite	**−451.58**	−444.42

Lower composite is better

Candidate 18 balances backbone, sequence, and modeled contact scores.

E Backbone contact refinement

Asp52 amide — Backbone O48

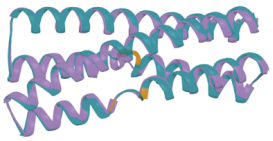


Target 1.87–2.05 Å

Late model ≈2.2 Å

The modeled H–O distance remains above target; the temporary hydrogen is omitted from export.

Evaluation part



Backbone RMSD 0.712 Å



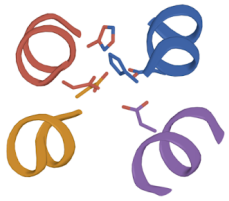
Motif RMSD 0.612 Å

Later redesign and refolding test motif and backbone recovery.

1BCF · Connect four helical fragments

motif · backbone · strands / Cα · refold · adjustment

A Motif constraints



8 residues each

- **A** EXXXXXXX — 1 fixed identity
- **B** XXXXEXXH — 2 fixed identities
- **C** XXEXXXXX — 1 fixed identity
- **D** XXXXEXXH — 2 fixed identities

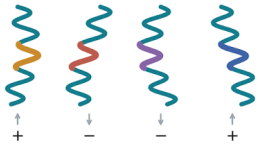
X = redesignable identity

N · O · S

Four fragments suggest a Due Ferri-like bundle; no metal is modeled.

B Supercoil geometry

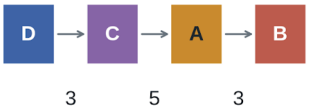
Supercoil model · no metal



Motif Cα fit: 0.318 Å

A curved helical model fits the motif and guides extension.

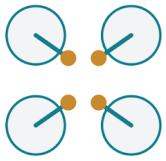
C Chain routing



125 residues total

Matched turns connect D–C–A–B after the sampling fix.

D Packing after refinement



Try Leu / Val shapes

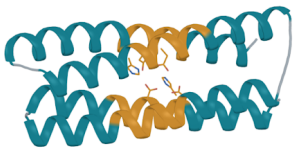
Count close contacts

Penalize overlaps

6 shuffled sweeps

After refinement reaches its limit, temporary Leu/Val models test packing; their atoms are omitted from export.

Evaluation part



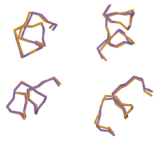
125 residues

4 motif pieces preserved

C–N links: 1.318–1.345 Å



Backbone RMSD 0.656 Å



Motif RMSD 0.328 Å

Later redesign and refolding test motif and backbone recovery.

5WN9 · Fit a second sulfur pair

motif — backbone — strands / Cα — refold — adjustment

A Motif constraints



A · 20 residues
XXXXXXCSNNPT
CWAICKXX
12 fixed identities

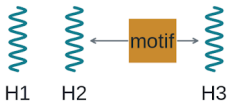
X = redesignable identity

N O S

Fixed SG pair: 2.051 Å

The supplied sulfur pair stays fixed; eight motif positions remain redesignable.

B Three-helix layout

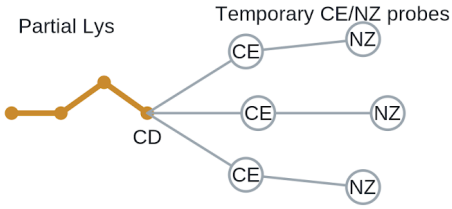


3,025 result files · no intermediate coordinates

6 Trp contacts vs 3

The selected layout has more tryptophan contacts despite a weaker combined score.

C Lys side-chain probes



8 branches; best internal score −36.156

Temporary CE/NZ probes filter refinement branches; these atoms are omitted from export.

D Sulfur-pair geometry



Geometry targets: 104° angles · |90°| torsion

A new cysteine is fitted toward fixed Cys17; displayed values are restraint targets.

E Submitted sulfur pairs

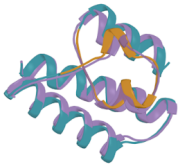


Fitted 46–59: 2.038 Å

Fixed 49–55: 2.051 Å

The submitted structure retains the fixed pair and includes the newly fitted cysteine.

Evaluation part



Backbone RMSD 1.113 Å



Motif RMSD 0.674 Å

SG pair submitted		refold
46–59	2.038	1.857
49–55	2.051	2.033

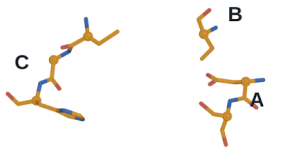
SG distances (Å) · diagnostic

Later refold sulfur distances are analyst diagnostics, not bond validation.

1B73 · Build around a Cys/His motif

motif · backbone · strands / Cα · refold · adjustment

A Motif constraints



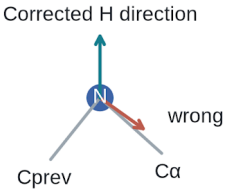
- **A · 2 residues** — DS
  2 fixed identities
- **B · 1 residue** — C
  1 fixed identity
- **C · 3 residues** — CXH
  2 fixed identities

X = redesignable identity

N · O · S

Three supplied fragments constrain the backbone and side-chain geometry.

B Virtual-hydrogen repair



16 rounded Cα positions

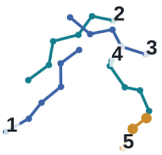
Exploratory test after repair



Rerun HB score: −41.17

Derivative agreement misses the wrong H direction; both formula and gradient signs are corrected.

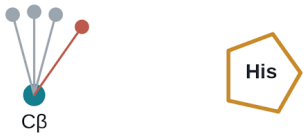
C Sheet-framework construction



5 separate partial chains

Five separate Cα chains show exploratory sheet geometry; final ancestry remains unresolved.

D Ser–His rotamer probe

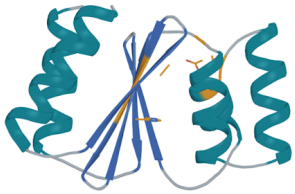


χ −55° probe · OG–NE2: 2.803 Å

∠Cβ–OG–NE2: 111.26°

One hypothetical Ser rotamer tests His proximity; its oxygen is omitted from export.

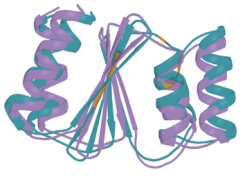
Evaluation part



125 residues

3 motif pieces preserved

C–N links: 1.327–1.331 Å



Backbone RMSD 1.770 Å



Motif RMSD 1.197 Å

Later redesign and refolding test motif and backbone recovery.

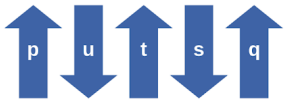
E.1.3 Topology-Conditioned Backbone Design

3.30.2310 · YaeB-like backbone construction

helix · β strand · loop

A Fold plan

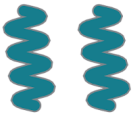
Antiparallel sheet



sheet order: p – u – t – s – q

packs against

Helix layer



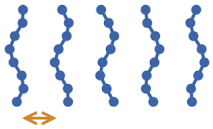
α1 + α2

The model chose a five-strand sheet with two helices from the YaeB/RelE-like taxonomy label.

B Coordinate construction

Analytic sheet

3.28 Å registry rise



adjacent spacing: 4.93 Å

Ideal helices



1.51 Å rise / residue
99.5° rotation / residue

Analytic targets define the parts; bond lengths, angles, and torsions form the connected chain.

C Fragment assembly

136,800 fragments processed → **136,541** retained by filters → **12 nearest** per matching query

2.1×10^{-5} relative gradient error

1147 → 183 objective • 100-iteration cap

Filter loop fragments, match connectors, and validate the gradient engine with a separate numerical test.

D Refinement and selection

480 calls polish job completed → **candidate 34** selected precursor → **6 local targets** 500-iteration cap

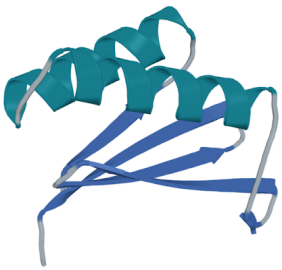
−72° to −55° precursor helix φ range

1.14 post-refinement repulsion

Internal diagnostics • the 500 value is a cap

Candidate 34 led the late comparison. Local torsion and hydrogen-bond restraints produced the final precursor.

E Submitted backbone and refold evaluation



87 residues · 348 backbone atoms

Chain A · N / Cα / C / O

- One continuous peptide
- C–N gate: 1.30–1.36 Å
- Logged radius of gyration: 12.075 Å

Selected refold

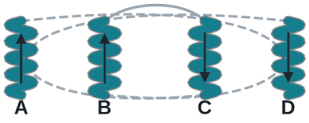
C.A.T match · RMSD 0.644 Å

1.20.1250 · Cytokine-like bundle construction

helix · β strand · loop

A Bundle topology

Up – up – down – down

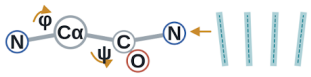


The model chose a cytokine-like bundle with two long returning loops. This is a working topology sketch.

B Coordinate construction

Grow atoms from bond frames

Soft helix targets



N / Cα / C / O + temporary Cβ

4 × 19-residue spans
1.46 Å rise / residue

A soft Cα sketch guides the bundle; internal coordinates produce the actual connected backbone.

C Candidate leaders

11 saved candidate leaders

run 1	**run 2**	**run 3**	**run 4**
run 5	**run 6**	**run 7**	**run 8** −30.97
run 9 −30.04	**run 10**	**run 11**	

190 candidates / run configured

staged shape → H bonds / sterics → packing

Eleven saved candidates were compared. The logs do not establish completion of every configured search candidate.

D Packing-rule revision

Common rescores · lower is better

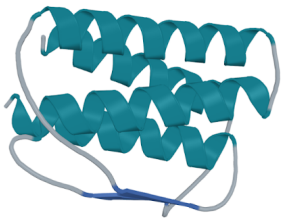


early run 8	**selected fin9**
3.86 Å	**4.38 Å**
nonadjacent Cα	nonadjacent Cα

Different candidates

Tighter steric limits changed the scoring rule. The two minima are unrounded measurements from different candidates.

E Submitted backbone and refold evaluation



116 residues · 464 backbone atoms

Chain A · N / Cα / C / O

- C–N readback: 1.3278–1.3298 Å
- Four long helices retained
- Final residue 116 and TER/END shown

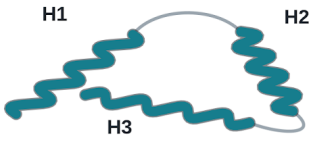
Selected refold

C.A.T match · RMSD 0.705 Å

1.10.12 · Three-helix bundle construction

helix β strand loop

A Three-helix hypothesis



Restrict the construction to three helices, then vary their relative placement and connecting loops.

B Peptide construction

NumPy / Torch builder



6 angles per residue

Ideal-chain test

O0–N4 = 2.989 Å

test geometry, not final PDB



Internal coordinates grow backbone atoms and a virtual Cβ. Restricted loop motifs interrupt the helical starting states.

C Packing and shape ranking

Packing leader · 19902 **Chosen · 7469**



H1–H3 cosine 0.92

axis-aware arrangement

packing 25.65 • shape −17.28

packing 20.92 • shape 12.67

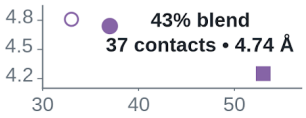
30,000 samples configured

The axis penalty demoted the nearly parallel bundle. The rollout chose 7469 despite its lower packing score.

D Geometric compromise

min Cα ≥2 res apart

loose: 33 contacts, 4.81 Å

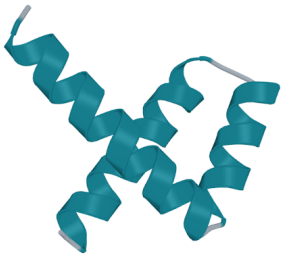


tight
53 contacts • 4.25 Å

contacts <8 Å, ≥5 residues apart

The blend occurs in angle space; points show unrounded measurements. The 43% choice is a manual geometric compromise.

E Submitted backbone and refold evaluation



56 residues · 224 backbone atoms

Chain A · N / Cα / C / O

- Three DSSP helices in the rounded PDB
- Unrounded: 37 Cα contacts <8 Å; ≥5 residues apart
- Unrounded: closest nonadjacent Cα 4.74 Å

Selected refold

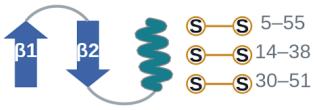
C.A.T match · RMSD 0.552 Å

4.10.410 · Kunitz-like backbone construction

helix β strand loop

A Kunitz-like topology

Core: hairpin + helix | **Virtual S–S restraints**



internal working labels

A Kunitz-like working model uses virtual disulfides; the final export retains a shorter neutral backbone slice.

B Diverse genetic search

Configured genetic operators

tournament ▸ **crossover** ▸ **diversity**

best of 1–6 draws | probability 0.45 | Cα pair distances

initial batch 320 / 136 later per pool — configured

Local minimization follows each proposal

Genetic proposals recombine torsions, then local optimization and pruning by Cα distances retain varied candidates.

C Checkpoint recovery

generation 23 → **h24 not found**

log observed → read arrived early

later in the same run

generation 72 → **h72 loaded**

log observed → refinement returned

The h24 read arrived before its checkpoint existed. Later logs and h72 supported the final refinement.

D Selection and geometry

Internal score • lower is better | **Working-model disulfide**

−7.84

selected end[0]

internal pair 14–38

2.03 Å

−5.48

next distinct signature

−87°

CB–S–S–CB torsion

sulfur atoms are auxiliary

Select end[0] and inspect its virtual disulfide geometry. The final export retains only a backbone slice.

E Submitted backbone and refold evaluation



46 residues · 184 backbone atoms

Chain A · N / Cα / C / O

- C–N readback: 1.328–1.330 Å
- Cα steps: 3.805–3.810 Å
- Logged radius of gyration: 10.026 Å

Selected refold

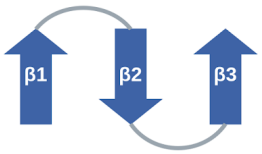
C.A.T match · RMSD 1.032 Å

2.20.70 · WW-like backbone construction

β strand loop

A WW-like topology

Antiparallel single sheet



Auxiliary packing



A WW-like three-strand sheet is the working plan, with temporary aromatic packing guides around its core.

B Search and diversity

8 searches → **GTT_49 • 4,550 complete**

launched post-loop log

Population diversity

distance-feature pruning

Other snapshots

completion not established

One GTT_49 run logged all 4,550 iterations; the other snapshots do not establish complete searches.

C Coordinate averaging

Robust alignment + coupled models

coordinate sets → **10 passes** robust target → **Two models** proper SVD fit

w / (1 + d² / 1.3)

configured robust weight

tie 3.34 • target 1.41

earlier polish diagnostics

Ten aligned averaging passes build a target; a coupled polish returns internal penalties under manual weights.

D Validation and finalization

Hydrogen-bond report

for i, j in pairs → **pairs.keys()**

unpacking error report succeeded

pol3 selected branch → **73% internal + 27% Cartesian** selected coordinate blend

Standalone DSSP format warning persisted

Using sparse matrix keys fixed the hydrogen-bond report; pol3 supplied the final coordinate blend.

E Submitted backbone and refold evaluation



31 residues · 124 backbone atoms

Chain A · N / Cα / C / O

- One continuous peptide: residues 1–31
- Cα steps: 3.799–3.837 Å
- Logged radius of gyration: 8.956 Å

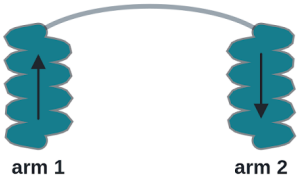
Selected refold

C.A.T match · RMSD 0.681 Å

1.10.287 · Helical hairpin construction

helix β strand loop

A Hairpin topology



Choose an antiparallel helical hairpin; a coiled coil target and a short turn constrain the geometry.

B Backbone construction

Internal coordinates



6 angles per residue
480-iteration cap

Surrogate objective

target fit

H bonds + sterics

packing + priors

Torch → L-BFGS-B

Internal coordinates fit the target while torsion priors, hydrogen bonds, and packing guide relaxation.

C Seed comparison

Completed seed sweep

21 calls	**seed 4**	**seed 15**
completed	selected	comparison
attraction	**−44.9**	**−38.5**

Internal terms • selection used several diagnostics

The 21-call sweep completed. Seed 4 was selected after comparing contacts, torsions, and internal energy terms.

D Turn repair

Residue 32 ψ	**Repulsion**
175.5° → 166.6°	**1.892 → 2.082**
measured before / rerun	increased
prior center 158° HB weight 0.91 → 1.19	internal diagnostic

global ω noise: noise[:,2] *= 0.88

After local prior/noise, global omega-noise ×0.88, and HB-weight changes, psi moved toward the target; repulsion rose.

E Submitted backbone and refold evaluation



63 residues · 252 backbone atoms

Chain A · N / Cα / C / O

- C–N readback: 1.325–1.336 Å
- Cα steps: 3.793–3.836 Å
- PCA helix-axis angle: 159.36°

Selected refold

C.A.T match · RMSD 1.068 Å

2.40.160 · Porin-like barrel construction

β strand loop

A Barrel family selection

Lipocalin	**Porin**	LamB
2.40.128	2.40.160	2.40.170

The taxonomy distinguishes nearby barrel families. The model chooses a porin interpretation and an OmpG-like plan.

B Analytic barrel target

Unrolled sheet Elliptical target



n = 14 strands
S = 16 shear

r (Å) = hypot(4.78n, 3.34S) / (2π)

The analytic target wraps a sheet into a barrel; both initial relaxation stages reached their iteration caps.

C Candidate comparison

Internal composite • lower is better

candidate 13	**candidate 37**
41.213	**41.328**

Candidate 37 was chosen although 13 had a lower score. The trace does not state the deciding reason.

D Local refinement

6	**6 HB**	**2,400 cap**
torsions targets changed	+ **weights** strengthened	→ reached

41.328 → 39.004	**0.973 Å**
same later composite	max Cα drift from b37

Targeted torsion and hydrogen-bond edits improved the same later composite; the optimizer again hit its cap.

E Submitted backbone and refold evaluation



187 residues · 748 backbone atoms

Chain A · N / Cα / C / O

- C–N readback: 1.322–1.338 Å
- Nonadjacent Cα minimum: 4.186 Å
- 315 contacts <8 Å; ≥12 residues apart

Selected refold

C.A.T match · RMSD 0.977 Å

3.10.20 · Ubiquitin-like backbone construction

helix β strand loop

A Contact hypothesis

Five-strand contact hypothesis



parallel β1–β5 pairing

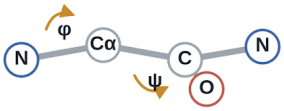
packs against



Translate the beta-grasp label into a mixed-sheet contact plan. This panel is a schematic of the imposed adjacency.

B Backbone relaxation

Frame-based atom builder



6 angles per residue

Four-stage test ran

1 · organize contacts

2 · add soft sterics

3 · tighten geometry

4 · score candidate

1,400 candidates configured

Internal coordinates grow the chain. A four-stage test ran; the larger 1,400-candidate launch lacks full completion evidence.

C Contact-distance checks

One early candidate: two violated guesses

O50 → N43 internal penalty 160.9 **5.54 Å**

O64 → N46 **5.26 Å**

internal penalty 131.2

2.95 Å target

These are internal optimizer labels. Both measured O–N distances were far from the configured target.

D Restraint revision

Revise contact restraints

~~**O50 → N43**~~

~~**O64 → N46**~~

196

saved finishing records

rank 2

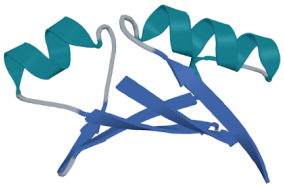
selected index 1

Tune caps and strand twist

then locally polished

The model read 196 finishing records and selected index 1 after revising restraints. Residual contact warnings remained.

E Submitted backbone and refold evaluation



81 residues · 324 backbone atoms

Chain A · N / Cα / C / O

- C–N readback: 1.3196–1.3418 Å
- Cα steps: 3.7855–3.8337 Å
- One model, one chain; backbone atoms present

Selected refold

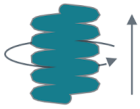
C.A.T match · RMSD 0.807 Å

1.10.472 · Cyclin-like backbone construction

helix · β strand · loop

A Ideal helix geometry

Cyclin-like plan • recalled context



98.74°

rotation / residue • test axis

1.500 Å

axial rise / residue • test

Test geometry; separate from final PDB

A recalled cyclin-like plan becomes ideal helices. A separate test checks the logged screw-axis construction.

B Coarse packing score

Excluded volume | **Hydrophobicity**



Cα / Cβ proxies | $0.15 + 0.4(h_i + h_j)$

soft overlap penalty | 5–9 Å contact window

Loop reach

≤16 Å

soft reachability bound depends on loop length

Recalled strings + manual overrides • surrogate score

Recalled sequence strings and manual overrides shape hydrophobic packing; these are internal objective terms.

C Search and clustering

270 × 40 → **10,800** → **exit 0**

fixed search loop | returned outcomes | process completed

Coarse Cα models
Kabsch RMSD <4 Å

2,877 / 3,000
first cluster • score-sorted subset

Correlated search steps • cluster size is not confidence

The search loop completed; Kabsch clustering summarizes the best coarse models, not independent confidence.

D Cap-contact repair

Working cap contact



Harmonic restraints

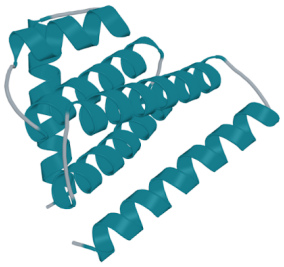
	Target	Measured
O–H	2.10 Å	**2.220 Å**
O–N	3.00 Å	**3.136 Å**

acceptor dot 0.101

low alignment diagnostic

After local restraints, both measured distances remain above their targets and the acceptor alignment diagnostic is low.

E Submitted backbone and refold evaluation



139 residues · 556 backbone atoms

Chain A · N / Cα / C / O

- C–N: 1.322–1.338 Å
- Cα steps: 3.792–3.838 Å
- DSSP: six long helix runs

Raw backbone audit: no topology recorded

Selected refold

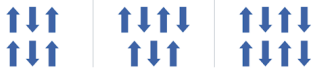
C.A.T match · RMSD 0.686 Å

2.60.420 · β-sandwich topology construction

β strand loop

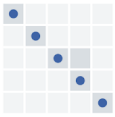
A Retained sheet layouts

62	148	164
6 strands	7 strands	8 strands



Restricted permutations generate two-sheet layouts. Counts cover only the explicitly imposed rules.

B Strand matching



$m = \exp(-1.5\,\Delta y^2)$

if sheet sign + strand direction agree

$S = \Sigma\, wR^4$

Initial 6-strand top score
0.220

Dynamic programming compares ordered strand slots under allowed flips and shifts; similarities are heuristic.

C Ranking sensitivity

Internal similarity; higher is better

521/4367 **0.152**

top score

7612/543 **0.132**

chosen later

Changing heuristic weights changes the ranking. The final seven-strand layout is not the top result of this rerun.

D Scaffold choice and construction



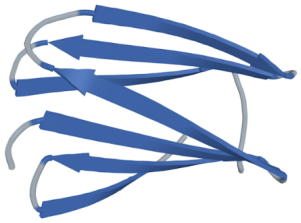
4.8 Å strand spacing

±5.05 Å sheet offset

Lateral order, not chain order; initial targets

The model selected 7612/543 as a structural hunch, then imposed a twisted scaffold and assigned hydrogen bonds.

E Submitted backbone and refold evaluation



74 residues · 296 backbone atoms

Chain A · N / Cα / C / O

Cα steps: 3.789–3.826 Å

C–N gate: 1.31–1.34 Å

Positive-φ ALA 53 and 66 remain

Raw backbone audit: 2.60.320

Selected refold

C.A.T match · RMSD 1.166 Å

1.10.510 · Kinase-like lobe construction

helix β strand loop

A Motif anchors

15 recalled strings • unverified



Only motif anchors shown

B Hydrophobicity prior

Approximate profile

176 columns

manual motif slices + gaps



profile mean working residue

$h = 0.65\,\bar{h}_{profile} + 0.35\,h_{work}$

plus per-position overrides

C Lobe geometry

Imposed lobe plan



Helices + motif torsion wells

Configured proxy distances

3.17 ±0.17 Å

N164/N165 → D220 proxy

6 ±1 Å

C-spine proxy contacts

Working labels • temporary sites

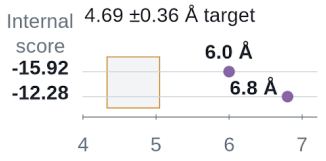
Fifteen recalled kinase-labeled strings are aligned around motif anchors; sequence and offset accuracy is unverified.

Motif-centered slices form an approximate profile; a 65/35 blend supplies a hand-built hydrophobicity prior.

Helices, torsion wells, and proxy contacts impose a kinase-lobe plan; they do not establish catalytic chemistry.

D Late-candidate audit

Arg165-CZ → phosphate proxy 197PH



Late pool only • final ancestry unresolved

Both printed late candidates miss the proxy band. Their relationship to the final saved PDB remains unresolved.

E Submitted backbone and refold evaluation



152 residues · 608 backbone atoms

Chain A · N / Cα / C / O

- One continuous peptide
- Finite backbone coordinates
- Chemical proxy atoms omitted

Raw backbone audit: target match

Selected refold

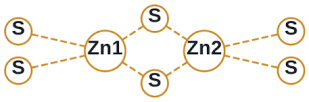
C.A.T match · RMSD 1.949 Å

4.10.240 · Gal4-like backbone construction

helix · β strand · loop

A Virtual zinc cluster

Virtual Zn_2–C_6 scaffold



Zn–Zn target: 3.19 Å

Zn–S targets: 2.31–2.36 Å

All Zn and S sites are temporary

Recalled Gal4-like context becomes two virtual Zn sites and six Cys proxies; the cluster is an imposed model.

B Repeated-motif search

repeat A **repeat B**

11-residue block copy • configured

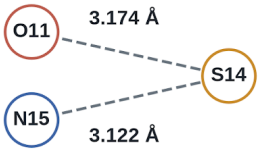
Soft restraints → **Full restraints**

initial relaxation → local relaxation

An 11-residue torsion block can be copied between repeats, followed by soft-to-full relaxation under configured rules.

C Cysteine contact

Working model



+89°

working Cys χ

Steric concern raised

No calibrated strain metric

Working display labels = internal indices + 10

An inspected Cys rotamer and sulfur contacts raise a steric concern; the labels belong to the working model.

D Virtual-geometry revision

Boundary changed → revised working model

Cys11 excluded from helix

Cys14 χ: −70°

2.27–2.40 Å

virtual Zn–S

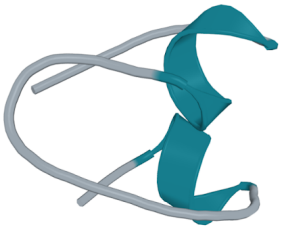
93.7°–118.0°

first-site S–Zn–S

Later revised candidate: internal repeat RMSD 0.31 Å

After the boundary change, check virtual Zn–S geometry and the internal repeat RMSD; temporary atoms are not submitted.

E Submitted backbone and refold evaluation



31 residues · 124 backbone atoms

Chain A · N / Cα / C / O

- ALA backbone; proxies omitted
- TER / END shown in output
- DSSP: two short helix runs

Raw backbone audit: no topology recorded

Selected refold

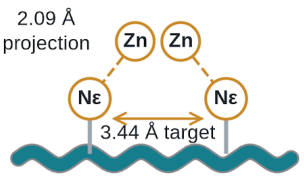
C.A.T match · RMSD 1.340 Å

3.30.2010 · Zincin-like scaffold construction

helix · β strand · loop

A Virtual histidine geometry

His pair · i and i+4



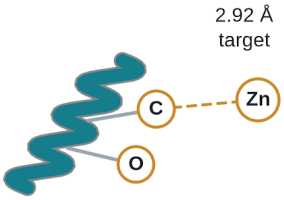
25	6.1
fits returned	best displayed cost

convergence flags not printed

Twenty-five His-rotamer fits returned; the 2.09 Å Zn projections are configured geometry, not measured metal bonds.

B Helix placement

Virtual Glu/Asp helix



11-residue helix

32 fits returned

1		**14.45**
2		14.88
3		15.18

internal cost · lower is better

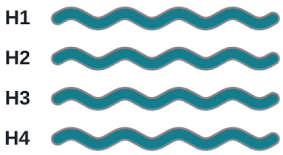
Thirty-two Glu/Asp fits returned. Their internal costs rank virtual-site poses; final-precursor linkage remains unverified.

C Scaffold assembly



Mixed sheet

packs with



chain: H1 – β1 – β2 – β3 – H2 – H3 – H4

Four helix blocks and three strand blocks form the working scaffold; restricted loop searches connect the parts.

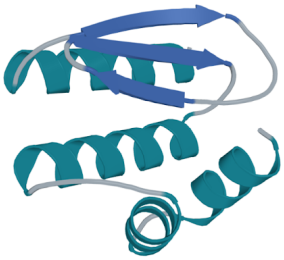
D Export and readback

rank[26] selected record → **97 residues** parser readback → **388 atoms** backbone only

1.3290–1.3309 Å	3.8069–3.8215 Å
C–N readback	Cα steps

Export rank[26], then check 97 residues and continuous backbone distances. The earlier site-pose ancestry is unresolved.

E Submitted backbone and refold evaluation



97 residues · 388 backbone atoms

Chain A · N / Cα / C / O

- C–N readback: 1.3290–1.3309 Å
- Final DSSP: 19 extended assignments
- Virtual Zn and sidechains omitted

Selected refold

C.A.T match · RMSD 1.555 Å

2.20.25 · Zinc ribbon backbone construction

β strand loop

A Virtual zinc targets

Four ligand-derived target sites



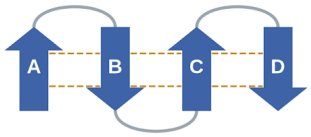
springs join projected sites

S–S 3.845 · S–N 3.56 · N–N 3.36 Å targets

Each ligand projects a virtual Zn position; springs join those positions and impose ligand-separation targets.

B Sheet-contact registers

Search sheet registers



AB · BC · CD contact families

4 Zn sites

configured

1 chain

full assembly

Search AB, BC, and CD hydrogen-bond registers, then combine them with the four virtual-site constraints in one chain.

C Candidate comparison

Same loop choice · logged alternatives

selected	lower energy	higher consensus
pick[2]	alternative	alternative
−43.26	**−47.93**	**−38.75**
internal energy	internal energy	internal energy
consensus 0.497	consensus 0.457	consensus 0.507

Trace choice: DSSP + TM-like consensus

Three logged candidates show energy/consensus tradeoffs. The trace cites DSSP and TM when choosing pick[2].

D Polishing and checks

pick[2] selected → **polish** steric 100 · Zn 62 → **saved sub** polished

0.0 Å

printed to 1 decimal

rounded site display



not exact coincidence

2,078 is an iteration cap; angles were not checked

Polished virtual-site deviations print as 0.0 Å to one decimal; exact coincidence and coordination angles remain unverified.

E Submitted backbone and refold evaluation



61 residues · 244 backbone atoms

Chain A · N / Cα / C / O

- C–N readback: 1.3281–1.3299 Å
- Four extended DSSP segments
- Virtual Zn and sidechains omitted

Selected refold

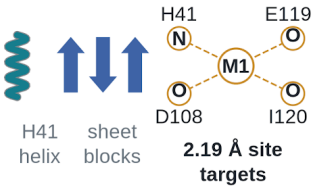
C.A.T match · RMSD 1.106 Å

3.40.91 · PA-like backbone construction

helix · β strand · loop

A Virtual metal sites

Early PA-like core | **Virtual metal 1**



An early PA-like core combines helices, sheet contacts, and virtual metal geometry; 2.19 Å is an imposed site distance.

B Discrete-case comparison

Internal penalties · lower is better

(79, NE, 1)		**(79, NE, 0)**	
metal	**1.64**	metal	**0.22**
sheet	**0.28**	sheet	**4.59**

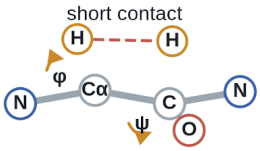
Case 0: O79–N108 = 4.23 Å

target 2.94 Å

Different discrete cases trade metal and sheet penalties; the 4.23 Å contact is measured against a 2.94 Å target.

C Short-contact repair

Add approximate H sites



≤12 overlaps
local distance targets

Proxy geometry → **2 repair passes**
cap: 1,260 iterations

7 seeds compared
logged results

Approximate hydrogen sites expose short contacts. Two repair passes add distance targets before seven seeds are compared.

D Selection and chain checks

Composite score · higher is better here

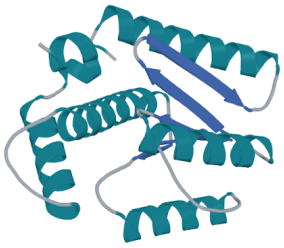
seed 72	**seed 35 · selected**
−233.821	**−230.714**

no top overlap reported	0.153 Å temporary HA overlap
	HA159–HA146 proxies
5 / 5	radial-closure passes simplified

heuristic chain check · not a formal knot proof

Seed 35 leads with a 0.153 Å temporary HA159–HA146 proxy overlap. Five radial-closure checks simplify the draft.

E Submitted backbone and refold evaluation



179 residues · 716 backbone atoms

Chain A · N / Cα / C / O

- C–N readback: 1.3281–1.3298 Å
- Closest nonadjacent Cα: 4.148 Å
- Virtual metals and ligands omitted

Selected refold

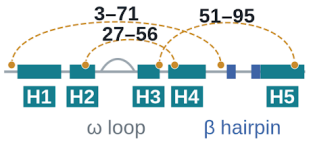
C.A.T match · RMSD 0.934 Å

1.10.239 · Sterol cavity backbone construction

helix · β strand · loop

A Family restraints

Recalled elicitin-like plan



Dashed arcs: disulfide hypotheses

Recalled elicitin-like features define a helical plan, an omega loop, a small sheet, and three disulfide connections.

B Temporary site models

Disulfide targets



1.97–2.08 Å

9 sulfur-rotamer pairs

Earlier candidate 1075



2.8 Å measured

nearest modeled ligand atom

Disulfide bands are targets; the 2.8 Å Tyr47–OH contact is measured in earlier ligand-modeled candidate 1075.

C Apo and holo comparison

25 candidates processed in each branch

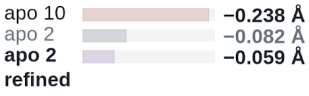
	apo · lr=0	holo · lr=0.4
index 10	**−297.92**	**−294.49**
index 2	**−296.28**	index 1 · leading separate objective

within-branch internal energy

Both branches ran on 25 candidates; compare energies within each branch because ligand weight changes the objective.

D Apo backbone export

Working backbone + Cβ margin · higher is better

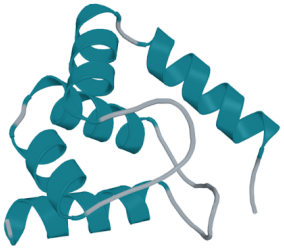


unrounded · margin = distance − threshold

Final script: apo[2] · lr=0 · backbone only

The final script refines apo[2] with lr=0 and omits the ligand; the earlier 2.8 Å contact does not describe this backbone.

E Submitted backbone and refold evaluation



98 residues · 392 backbone atoms

Chain A · N / Cα / C / O

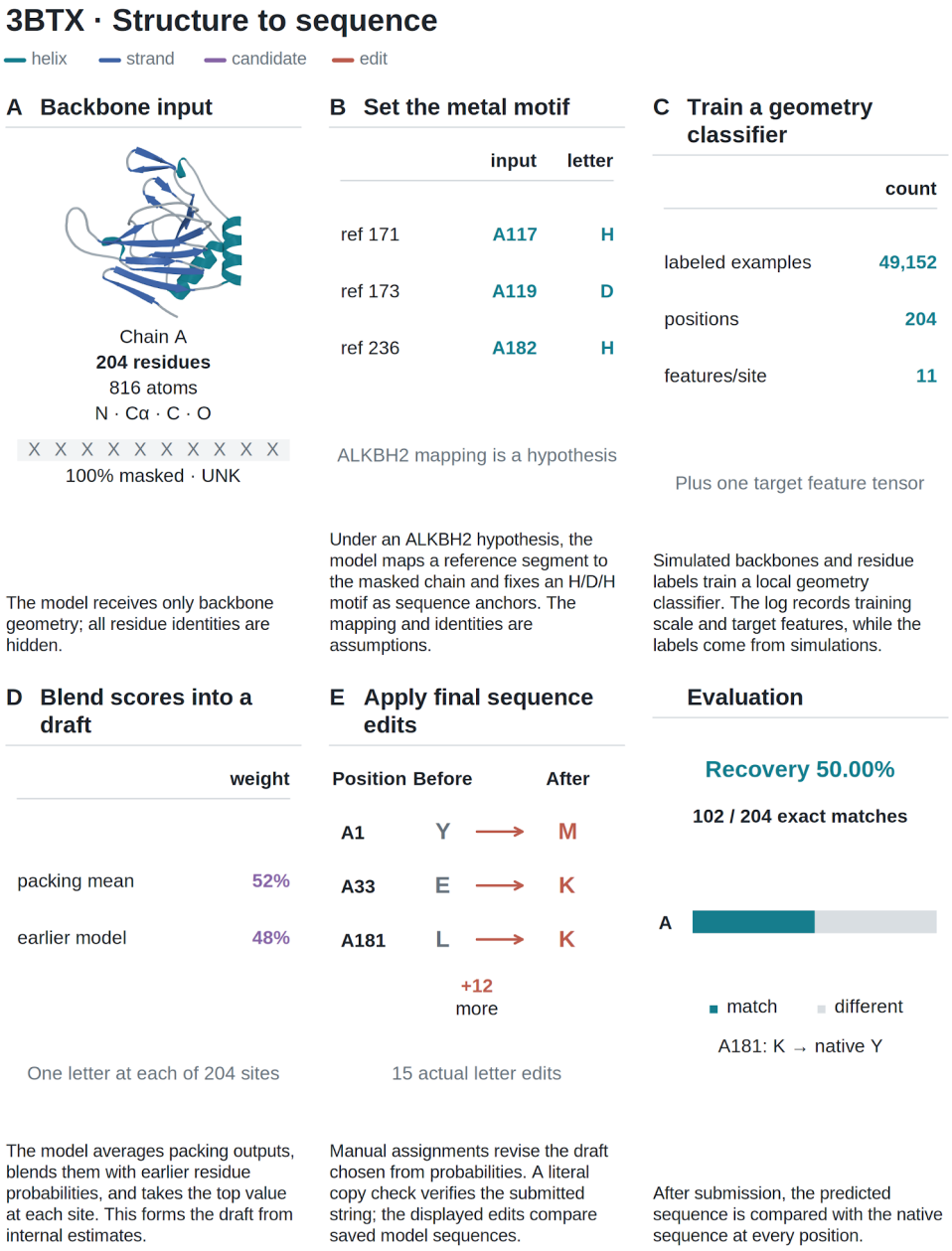
- Final branch: apo[2] • lr=0
- C–N readback: 1.3291–1.3310 Å
- Temporary ligand and sidechains omitted

Selected refold

C.A.T match · RMSD 1.069 Å

E.2 Sequence Design

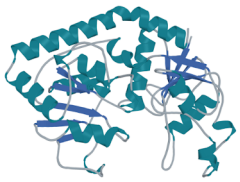
E.2.1 Protein Inverse Folding



1M5R · Structure to sequence

helix strand candidate edit

A Backbone input



Chain A
351 residues
1,404 atoms
N · Cα · C · O

X X X X X X X X X X
100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Summarize backbone geometry

	Residues
Helix	147
Strand	64
Other	140

Temporary ALA residue labels

On a temporary copy labeled ALA, DSSP summarizes helix, strand, and other regions. Coordinates stay unchanged; the longest adjacent C–N link is 1.349 Å.

C Build the baseline sequence

	Setting
Bonus mean	84%
Earlier mean	16%
Arrays per mean	8
Draft length	351

Then residue and charge biases

Two saved array means are blended, including one with model letter bonuses. Composition, charge, and helix biases then guide the top letter selected at each site.

D Apply a hypothetical DNA prior

	Value
Axes evaluated	8,000
Min. penalty	60.643
Letter changes	7

Hypothetical line geometry

A line model samples hypothetical DNA axes and biases charged letters by contact geometry. The displayed penalty and sequence changes quantify this assumed geometric prior.

E Finish the sequence

Position	Before		After
A115	R	→	Q
A258	K	→	D

Final literal edits

Two literal overrides modify the draft after the DNA prior. A direct string comparison verifies nine final differences from the baseline, including the seven earlier changes.

Evaluation

Recovery 52.14%

183 / 351 exact matches

A



match different

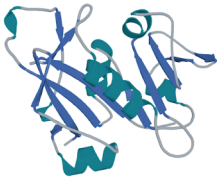
A25: L → native Y

After submission, the predicted sequence is compared with the native sequence at every position.

4HOF · Structure to sequence

helix · strand · candidate · edit

A Backbone input



Chain A
192 residues
768 atoms
N · Cα · C · O

X X X X X X X X X X
100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Supply a candidate template

DHFR hypothesis

M	V	G	S	L	N	C
I	V	A	V	S	Q	N

187 stored letters for 192 sites

Under a DHFR hypothesis, the model supplies a candidate sequence shorter than the masked backbone. Its letters guide later scoring; the family and species labels remain hypotheses.

C Generate side chain candidates

	Count
Conformers	98,253
Max per site	641

Candidate side chain coordinates

A local rotamer library generates candidate side chain coordinates on the backbone. The bank defines candidates for later scoring; a full design search remains unverified.

D Select a scored draft

	Value
Model 1 weight	0.42
Model 2 weight	0.58
Draft score	108.9

Sum of top site probabilities

Two internal residue models are blended with template priors, then the top letter at each site forms a complete draft. The displayed score sums those internal top probabilities.

E Apply final overrides

Position	Before		After
A1	T	→	M
A68	A	→	P
A144	E	→	Q

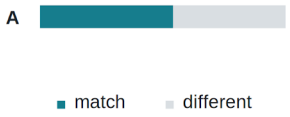
3 examples; 9 total changes

Manual assignments revise the scored draft, with one assignment leaving its letter unchanged. The copy check confirms the final string; the visual shows selected changes.

Evaluation

Recovery 54.17%

104 / 192 exact matches



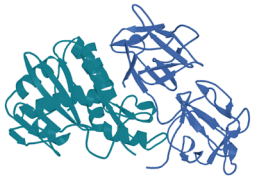
A68: P matches native

After submission, the predicted sequence is compared with the native sequence at every position.

1YF8 · Structure to sequence

Chain A Chain B candidate edit

A Backbone input



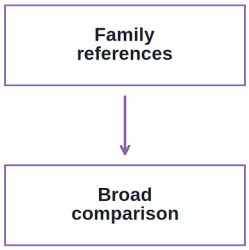
Two chains

Backbone only

Sequence identities hidden

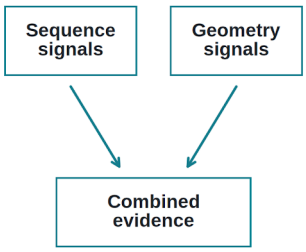
Two protein backbones are supplied with sequence identities hidden.

B Family comparison



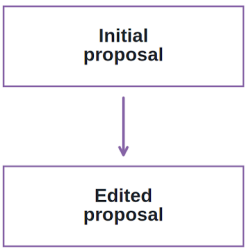
The model compares family references it supplied to guide broad sequence choices.

C Sequence and geometry



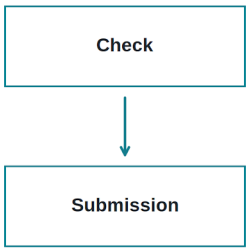
Sequence comparisons and backbone geometry are considered together to support the proposal.

D Edited proposal



The model combines its selected evidence into a proposal and revises that proposal before submission.

E Consistency check

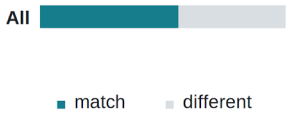


The model checks consistency of the edited proposal and submits the final result.

Evaluation

Recovery 56.36%

279 / 495 exact matches



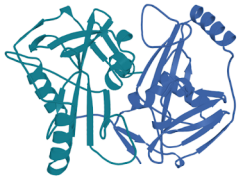
The saved aggregate recovery counts exact matches across the two supplied chains.

Detailed toxin sequence data are omitted.

5R66 · Structure to sequence

Chain A Chain B candidate edit

A Backbone input



Chain	Residues
A	197
B	197

1,576 atoms

N · Cα · C · O

X X X X X X X X X X

100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Compare the backbone copies

	Max error (Å)
A→B Cα	5.22e-4

Unconstrained affine coordinate fit

An affine map fits chain A Cα coordinates to chain B. The component error measures agreement under this flexible transformation; stretch and shear are allowed.

C Build a contact model

	Count
Conformers	74,909
Residue edges	2,022
Pair entries	269.15M

Returned table sizes

Custom code scores clashes, attraction, charges, and directional hydrogen bonds across both copies. The returned table sizes describe candidate coverage, not physical accuracy.

D Form one shared draft

	Value
Saved variants	10
Shared sites	197
Mean top p	0.474

Top letter selected at each site

Saved packing variants are averaged into residue probabilities at shared sites. The top letter at each site forms one draft; the mean top probability summarizes internal confidence.

E Join the prefix and draft

Position	Before	After
34	W →	F
69	V →	L
112	V →	L

Then copied to A/B

The model joins a literal prefix of 30 letters to 167 scored draft letters, then makes three edits. The result is copied to both chains; block lengths come from saved strings.

Evaluation

Recovery 58.38%

230 / 394 exact matches



match different

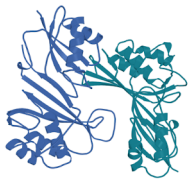
A34: F matches native

After submission, predicted and native sequences are compared. Recovery pools exact matches across all chains.

1VS3 · Structure to sequence

Chain A Chain B candidate edit

A Backbone input



Chain	Residues
A	249
B	249

1,992 atoms

N · Cα · C · O

X X X X X X X X X X

100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Compare the backbone copies

	value
Cα pairs	249
RMSD	0.567 Å

Rigid fit of input coordinates

A rigid Cα fit compares the supplied chains, and the same command returns tables of secondary structure. The result quantifies geometric agreement; sequence identity remains a model choice.

C Form a draft from weights

Returned candidate: 249 letters

K V R I L L A I

Top identity at each site

A custom sampler selects the top amino acid at each site from sampled weights. The fragment comes from its returned candidate, after a division warning in the calculation.

D Add an RNA contact prior

	coefficient
Arg/Lys	+1.43
His	+0.47
Asp/Glu	−0.77

Scaled by (contact score − 0.24)

A probe labeled for RNA samples placements and biases charged residues by contact score, then seven weight groups are blended. The probe and coefficients are model assumptions.

E Reweight and finish the sequence

Position	Before		After
4	V	→	F
13	A	→	S
245	G	→	E

249 letters copied to A/B

An exponential adjustment renormalizes residue weights before three explicit replacements. The final sequence is copied to both chains, so the displayed edits affect one shared proposal.

Evaluation

Recovery 60.64%

302 / 498 exact matches



match different

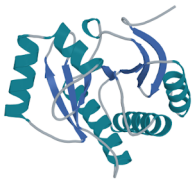
A13: S → native T

After submission, predicted and native sequences are compared. Recovery pools exact matches across all chains.

6FY5 · Structure to sequence

helix strand candidate edit

A Backbone input

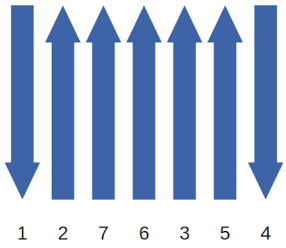


Chain A
198 residues
792 atoms
N · Cα · C · O

X X X X X X X X X X
100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Model topology



The model reads the secondary structure table and proposes this sheet topology.

C Residue preferences

Residue	Internal frequency
L17	93%
T35	67%
L51	71%

A custom model estimates residue preferences from packing, hydrogen bonds, and solvation. An import repair lets the calculation finish.

D Candidate alignment

205
recalled letters

198
modeled sites

Internal score 124.74 · gap cost −6

A macrodomain hypothesis prompts a recalled candidate, which the model aligns to the backbone sites.

E Hybrid sequence

40 model-entered 158 recalled

Example · positions 139–141

Before S I G
After P I P
I140 retained

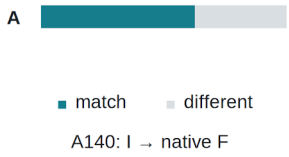
20
total changes

The model joins the two sequence sources, then edits the hybrid. Comparing the full saved sequences gives the total change count.

Evaluation

Recovery 62.63%

124 / 198 exact matches

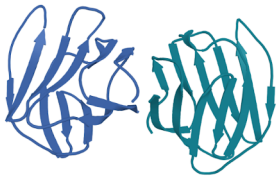


After submission, the predicted sequence is compared with the native sequence at every position.

6A64 · Structure to sequence

Chain A Chain B candidate edit

A Backbone input



Chain	Residues
A	138
B	138

1,104 atoms

N · Cα · C · O

X X X X X X X X X X

100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Crop the stored template

142 → 138

stored letters reference sites

Template positions 2–139

The model enters a candidate and crops it to match the backbone length. The retained letters supply reference weights for later scoring; the template identity remains a hypothesis.

C Fix selected residue identities

	sites
Gly	8
Pro	8
Cys	2

Imposed residue identities

Backbone angles modify weights from the template, while selected Gly, Pro, and Cys positions are fixed. The table records imposed identity constraints for later scoring.

D Blend and adjust the weights

	value
arrays	10
boosted sites	8
factor	×50

Then adjust mass/charge weights

Saved arrays are averaged and selected choices boosted before mass and charge reweighting. The returned candidate has 139 letters; a log warning appears and solver success is unchecked.

E Edit and trim the sequence

Position	Before	After
34	C	S
38	A	D
75	E	L

138 letters copied to A/B

Two literal replacements change three letters, then the leading M is removed to match the backbone length. The resulting sequence is copied to both chains as an explicit model choice.

Evaluation

Recovery 64.49%

178 / 276 exact matches



match different

A75: L → native E

After submission, predicted and native sequences are compared. Recovery pools exact matches across all chains.

4LX8 · Structure to sequence

helix · strand · candidate · edit

A Backbone input



Chain A
124 residues
496 atoms
N · Cα · C · O

X X X X X X X X X X
100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Candidate sequence source

candidate start

M	A	D	K	E	L	K
F	L	V	V	D	D	F

candidate end

L	E	E	K	L	N	K
I	F	E	K	L	G	M

129 letters entered by the model

The model scans backbone torsions and secondary structure, then supplies a CheY candidate. The fragments show the entered sequence; family and species remain hypotheses.

C Match the backbone length

129 → 124

stored reference · working candidate

Reference positions 2–125

The working sequence is a crop of the stored reference, omitting its initial M and terminal KLGM. Comparing saved strings confirms that the same crop is later submitted.

D Compare sampled rotamers

	χ1	score
sample 1	−82°	0.0
sample 2	+67°	41.8

Candidate Y51 = input A50

The model compares sampled Tyr side chain orientations at one candidate site. Relative internal scores rank those conformers; they do not test whether Tyr is the correct residue.

E Inspect backbone geometry

	distance
A21 C–A22 N	1.167 Å

Supplied backbone measurement

A final scan measures the peptide link in the supplied coordinates. The sequence remains the earlier reference crop, so this geometry check does not establish exact identity.

Evaluation

Recovery 66.94%

83 / 124 exact matches



match · different

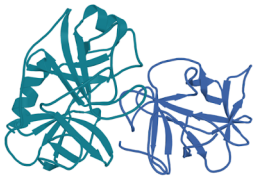
A50: Y → native F

After submission, the predicted sequence is compared with the native sequence at every position.

7JQV · Structure to sequence

Chain A Chain B candidate edit

A Backbone input



Chain	Residues
A	223
B	166

1,556 atoms

N · Cα · C · O

X X X X X X X X X X

100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Build chain A from two sources

152 → 71

remembered prefix constructed tail

Chain A: 223 letters

A remembered prefix supplies most of chain A; motif and sampled choices supply the tail. The split shows the direct contribution of recall, while the family label remains a hypothesis.

C Align the inhibitor reference

	ref. pos.	prob.
B7	5	0.99
B65	63	1.00
B155	168	1.00

B65→63 was an imposed anchor

The model aligns a remembered Kunitz candidate of 180 letters to B's 166 sites. An imposed anchor shapes the displayed probabilities; they are internal alignment outputs.

D Check proline preferences

	bond	shape
B63	+2.3	0.0
B65	-1.2	-2.2

Rounded custom Pro components

The model compares these local terms with Pro frequencies of 0.923 at B63 and 7.8e−21 at B65. Disfavoring Pro does not choose the correct alternative residue.

E Assemble the submitted pair

before A153–159

L P T Y T P S

final A153–159

D P L Y H P S

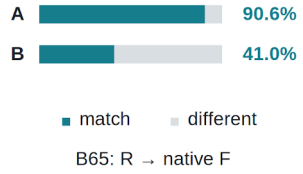
One of three A motif replacements

Three motif replacements in A and 15 edits in B produce the submitted pair. Final checks confirm lengths 223/166 and only allowed amino acid letters; the tiles show one A example.

Evaluation

Recovery 69.41%

270 / 389 exact matches



B65: R → native F

After submission, predicted and native sequences are compared. Recovery pools exact matches across all chains.

1G9Y · Structure to sequence

Chain A Chain B candidate edit

A Backbone input



Chain	Residues
A	152
B	152

1,216 atoms

N · Cα · C · O

X X X X X X X X X X

100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Supply sequence priors

	Count
Prefix letters	34
Other labels	17
Total priors	51

I-CreI identity is a hypothesis

Under an I-CreI hypothesis, the model supplies a prefix and additional residue labels as priors. The table counts model assumptions rather than letters measured from the backbone.

C Guard against reflections

−1 → +1

if det(R) < 0 after vector flip

Conditional guard; then rerun

The alignment code adds a conditional guard against reflected coordinate frames, then reruns packing and fitting. The returned output does not show whether the guard was triggered.

D Sample a draft under priors

	Value
Input 35 K	83%
Input 36 H	96%
Draft letters	152

Internal frequencies under priors

A hypothetical DNA axis and explicit residue priors shape the sampling pass. The displayed frequencies and draft length are returned model outputs, not final native recovery.

E Edit and copy the draft

Position	Before		After
A71	D	→	N
A89	E	→	Q
A125	C	→	A

3 examples from 8 edits

Eight literal assignments revise the draft, which is copied to both chains. The JSON copy check returns true; displayed input positions are one below the model's numbered tags.

Evaluation

Recovery 69.74%

212 / 304 exact matches



match different

A125: A → native C

After submission, predicted and native sequences are compared. Recovery pools exact matches across all chains.

1TVQ · Structure to sequence

helix · strand · candidate · edit

A Backbone input



Chain A
125 residues
500 atoms
N · Cα · C · O

X X X X X X X X X X
100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Fit hypothetical cysteines

	Value
Cα gap (Å)	5.638
Virtual Cβ (Å)	4.268
Best fit score	8.779

10 starts; C80/C91 hypothesis

The model fits sulfur positions for hypothetical cysteines. The score sums squared scaled errors, assessing the constructed geometry without confirming a disulfide bond.

C Turn templates into priors

Template 1

S	F	S	G	K	Y	Q
L	Q	S	Q	E	N	F

Template 2

A	F	T	G	K	Y	E
I	E	S	E	K	N	Y

Both trimmed to 125 sites

Under a FABP hypothesis, two stored candidates are trimmed to the backbone length. BLOSUM80 and a matrix square root convert those candidate letters into residue priors.

D Build a constrained draft

	Value
Variants	5
C80/C91 fixed	4
Top prob. sum	81.06

125 letters from top site scores

Packing variants are averaged, then the top score at each site forms the draft. The cysteine constraints are imposed; the probability sum is an internal model summary.

E Apply final sequence edits

Position	Before		After
A10	S	→	A
A46	H	→	D
A61	E	→	K

3 examples from 6 edits

Six literal assignments revise the draft before submission. The final sequence keeps the required length, and an exact JSON copy check returns true; the table shows three examples.

Evaluation

Recovery 72.80%

91 / 125 exact matches



match · different

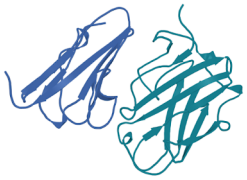
A91: C → native T

After submission, the predicted sequence is compared with the native sequence at every position.

1P6A · Structure to sequence

Chain A Chain B candidate edit

A Backbone input



Chain	Residues
A	185
B	124

1,236 atoms

N · Cα · C · O

X X X X X X X X X X

100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Build B from stored letters

Stored prefix

I T T P E E M

After adding S

S I T T P E E M

Chain B: 123 → 124 letters

The model prefixes S to a stored sequence block to build chain B. The tiles show the literal operation, making clear that these candidate letters come from the entered block.

C Set template priors for A

	size
ad5 string	182
ad3 map	185

String letters / map positions

For chain A, two model-supplied templates set starting letter preferences. One is a sequence string and the other a position map with gaps and unknowns; their names remain hypotheses.

D Average weights to build A

	weight
A7 T	99%
A9 W	100%
A18 C	98%

Mean of three saved outputs

Saved model outputs are averaged, then the top amino acid at each of 185 sites forms chain A's draft. The displayed weights include template priors and are not native recovery.

E Edit the submitted pair

Position	Before		After
A4	I	→	N
A85	E	→	Q
A161	V	→	W

+8 more

Final lengths: A185 / B124

Eleven replacements revise chain A while B keeps the prefixed stored block. The final JSON passes an exact copy check; the visual shows three examples and the final chain lengths.

Evaluation

Recovery 76.38%

236 / 309 exact matches



match different

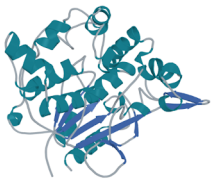
A85: Q matches native

After submission, predicted and native sequences are compared. Recovery pools exact matches across all chains.

4FMP · Structure to sequence

helix · strand · candidate · edit

A Backbone input



Chain A
381 residues
1,524 atoms
N · Cα · C · O

X X X X X X X X X X
100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Fit hypothetical metal sites

	Max error (Å)
Pattern 1	0.062
Pattern 2	0.885
Pattern 3	0.689

Four virtual Cβ distance targets

The model fits three hypothetical metal centers to virtual Cβ distance targets. Residuals measure agreement with those imposed distances, not confirmation of an ion or binding site.

C Store a candidate sequence

Sequence block

A	S	P	R	A	N	D
A	P	I	V	L	L	H

192 stored letters before trimming

The model enters a candidate sequence block, whose opening letters are shown. This stored block supplies the later prefix, but its protein identity is not independently established.

D Build a scored draft

8 → 381

saved sampling variants · top letters in draft

Prefix bias + packing energies

Saved sampling variants are averaged with a bias from the stored prefix and custom packing energies. Top site scores form the draft; they do not establish native sequence identity.

E Join and edit sequence blocks

	Letters
Stored prefix	188
Literal bridge	11
Scored tail	182

381 letters before 7 edits

The model joins the three blocks, then applies seven edits, including A207 Y→W. The JSON copy check returns true; input positions are four below the model's numbered tags.

Evaluation

Recovery 77.95%

297 / 381 exact matches



match · different

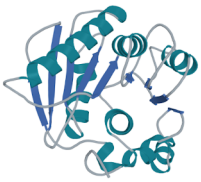
A207: W matches native

After submission, the predicted sequence is compared with the native sequence at every position.

7DZU · Structure to sequence

helix strand candidate edit

A Backbone input



Chain A
268 residues
1,072 atoms
N · Cα · C · O

X X X X X X X X X X
100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Supply a sequence template

Template start

T	N	P	Y	A	R	G
P	N	P	T	A	A	S

Motif region

M	G	W	S	M	G	G
G	G	S	L	I	S	A

Model-entered sequence fragments

The model enters a template under a PET hydrolase hypothesis. These stored fragments supply candidate letters for later local checks and sequence assembly.

C Compare local preferences

	Probability
A4 · A	0.50
A4 · Y	0.014
A49 · R	0.30
A49 · T	0.036

Template letters: Y4 and T49

Internal probabilities favor alternatives to the template letters at these sites, exposing a disagreement for later assembly; the calculation also emitted a division warning.

D Test imposed cysteine pairs

	Min. penalty
A174–A210	0.706
A244–A264	0.138

Imposed pairs; internal penalties

The code scores two imposed cysteine pairs; the finite minima quantify geometry under that assumption, while the recorded division warning remains a limitation.

E Join template and tail

	Residues
Template prefix	157
Literal patch	7
Template block	80
Literal tail	24

Returned length: 268

The final sequence joins template blocks with a literal patch and tail; checks of the saved strings confirm the assembled length and retention of the two earlier template choices.

Evaluation

Recovery 79.48%

213 / 268 exact matches



match different

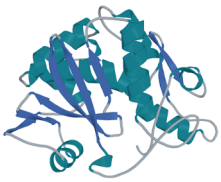
A5: A → native Q

After submission, the predicted sequence is compared with the native sequence at every position.

5CSO · Structure to sequence

helix · strand · candidate · edit

A Backbone input



Chain A
246 residues
984 atoms
N · Cα · C · O

X X X X X X X X X X

100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Build partial sequence hints

91 → 61 → 94

recalled prefix · unknown middle · homolog tail

246 positions in the hint string

The model combines two stored reference blocks around an unknown middle, creating a partial hint string for later sampling; the family labels remain hypotheses.

C Bias catalytic candidates

	letter	mult.
A70	Y	3×
A111	Y	3×
A160	E	3×
A163	R	3×

Prior match rate: 0.99999

The code increases sampling and prior preference for the displayed catalytic candidates; these configured biases guide packing but do not establish the residue identities.

D Average frequencies at A50

	frequency
V	55%
A	14%
T	10%
S	8%

Consensus from five outputs

Averaging saved internal frequency arrays produces a full candidate, while the displayed site disagrees with the remembered hint and leaves a final model choice.

E Finalize the two edits

Position	Before	After
A50	V	M
A92	A	S

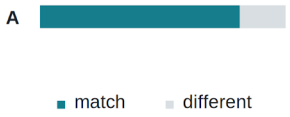
Two edits · 246 letters

The model restores the remembered letter at one site and makes a second motif edit, producing the final response string; both remain model choices before grading.

Evaluation

Recovery 81.30%

200 / 246 exact matches



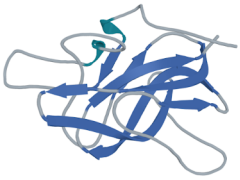
A50: M matches native

After submission, the predicted sequence is compared with the native sequence at every position.

3BN6 · Structure to sequence

helix · strand · candidate · edit

A Backbone input



Chain A
158 residues
632 atoms
N · Cα · C · O

X X X X X X X X X X
100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Enter and edit a candidate

Position	Before	After
A110	G	D

Candidate length: 158

The model begins with a complete entered sequence and makes the displayed substitution, so most candidate letters are supplied before the physical checks begin.

C Set terminal cysteine states

Setting	
S–S distance	2.03 Å
States per end	1

Imposed before packing

The code selects and fixes one rotamer at each terminus using the displayed sulfur distance target, carrying imposed cysteine geometry into the packing trials.

D Relax two packed models

	Etot
Trial 1 start	15,219
Trial 1 end	1,564
Trial 2 start	13,472
Trial 2 end	1,642

Rounded internal energies

Each packed model is minimized and returns a lower internal total energy. This checks the relaxation procedure; it does not establish convergence or final recovery.

E Submit the retained candidate

Position	Before	After
A2	A	T
A3	N	E
A11	S	T

Final template differences

The final sequence retains 155 of 158 letters from the stored edited candidate; the displayed differences are an analyst comparison of saved model strings.

Evaluation

Recovery 83.54%

132 / 158 exact matches



match · different

A9: N → native D

After submission, the predicted sequence is compared with the native sequence at every position.

6Z5Q · Structure to sequence

Chain A Chain B Chain C candidate edit

A Backbone input



Chain	Residues
A	90
B	90
C	90

1,080 atoms

N · Cα · C · O

X X X X X X X X X X

100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Set fixed and variable sites

Positions	
Fixed prefix	57
Other fixed	7
Variable	26

Baseline length: 90

The model joins a stored prefix and literal tail blocks, then fixes the displayed letters for packing. The other sites vary; fixed identities are configured assumptions.

C Repair the packing run

Result	
Initial run	KeyError
Rerun indices	90
Relax score	79.316

Returned internal outputs

A rotamer lookup error stops packing. After repair, the tool returns a state list and a later internal relaxation score; these outputs do not establish convergence.

D Check repeats and properties

Value	
Repeat match	24 / 45
Mass (kDa)	9.728
pI	6.04

Fixed gaps; calculated properties

The model compares repeats using fixed gaps, then calculates mass and pI from the candidate sequence; these checks describe the proposal before submission.

E Submit one shared sequence

Residues	
Chain A	90
Chain B	90
Chain C	90

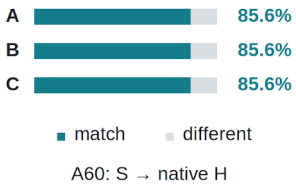
Identical sequence in all chains

The model copies the same final sequence to all three chains and retains every configured fixed letter, keeping the submission consistent with the fixed part of its baseline.

Evaluation

Recovery 85.56%

231 / 270 exact matches



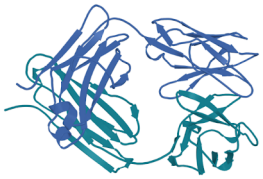
A60: S → native H

After submission, predicted and native sequences are compared. Recovery pools exact matches across all chains.

6U1T · Structure to sequence

Chain A Chain B candidate edit

A Backbone input



Chain	Residues
A	221
B	214

1,740 atoms

N · Cα · C · O

X X X X X X X X X X

100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Test cysteine placements

	Fit score
B134–B194	1.69
A145–A200	2.40
B23–B88	2.63

Selected pairs; lower score is better

The scan tests hypothetical cysteine placements and returns selected geometric fit scores. Lower values mean closer agreement within this model; bonds remain unconfirmed.

C Compare framework rankings

	Score
List 1 · vk12b	−88.8
List 1 · vk632	−89.6
List 2 · vk632	−92.3
List 2 · vk12b	−93.0

Scores rank candidates within each list

One calibration call returns two internal rankings with different leaders. The displayed comparison shows that the preferred framework changes between the reported lists.

D Assemble the chain sequences

Position	Before	After
B91	H	Y

Final literal replacement

The model edits a supplied sequence block and appends stored tails for both chains. The displayed replacement is a literal choice; returned lengths check the resulting assembly.

E Check the submitted copy

	Residues
Chain A	221
Chain B	214

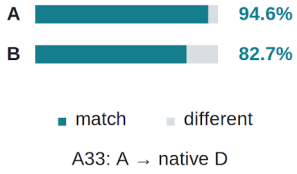
Printed strings equal the final JSON

The printed chain lengths and sequences match the final JSON, checking that the assembled strings were copied consistently before the separate native evaluation.

Evaluation

Recovery 88.74%

386 / 435 exact matches



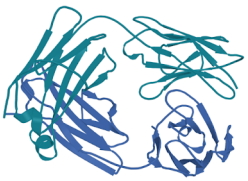
A33: A → native D

After submission, predicted and native sequences are compared. Recovery pools exact matches across all chains.

1NGW · Structure to sequence

Chain A Chain B candidate edit

A Backbone input



Chain	Residues
A	213
B	216

1,716 atoms

N · Cα · C · O

X X X X X X X X X X

100% masked · UNK

The model receives only backbone geometry; all residue identities are hidden.

B Measure terminal geometry

	Distance
A213–B216	7.34 Å
B129–B216	8.38 Å

Computed virtual Cβ positions

The code computes virtual Cβ positions from the backbone and measures terminal separations, checking local geometry while residue identities remain unassigned.

C Build a conformer bank

	Count
Conformers	59,884
Atom slots	14
Fields	4

Candidate residues; padded array

The code builds candidate residue shapes from 20 amino acid templates and stores them in a padded array; the displayed dimensions describe this coordinate bank.

D Fit weights to candidate labels

	Log loss
Initial	2.275
Fitted	2.169

Internal candidate labels

The model fits eight score weights to labels from a complete candidate pair. The lower log loss measures fit to those assumed letters; native recovery is evaluated separately.

E Join blocks and final edits

Position	Before		After
A11	L	→	M
B72	A	→	V

Final literal replacements

Stored candidate blocks and tails supply both chains, followed by the displayed literal replacements. Saved outputs verify A/B lengths 213/216 and the edit positions.

Evaluation

Recovery 89.28%

383 / 429 exact matches

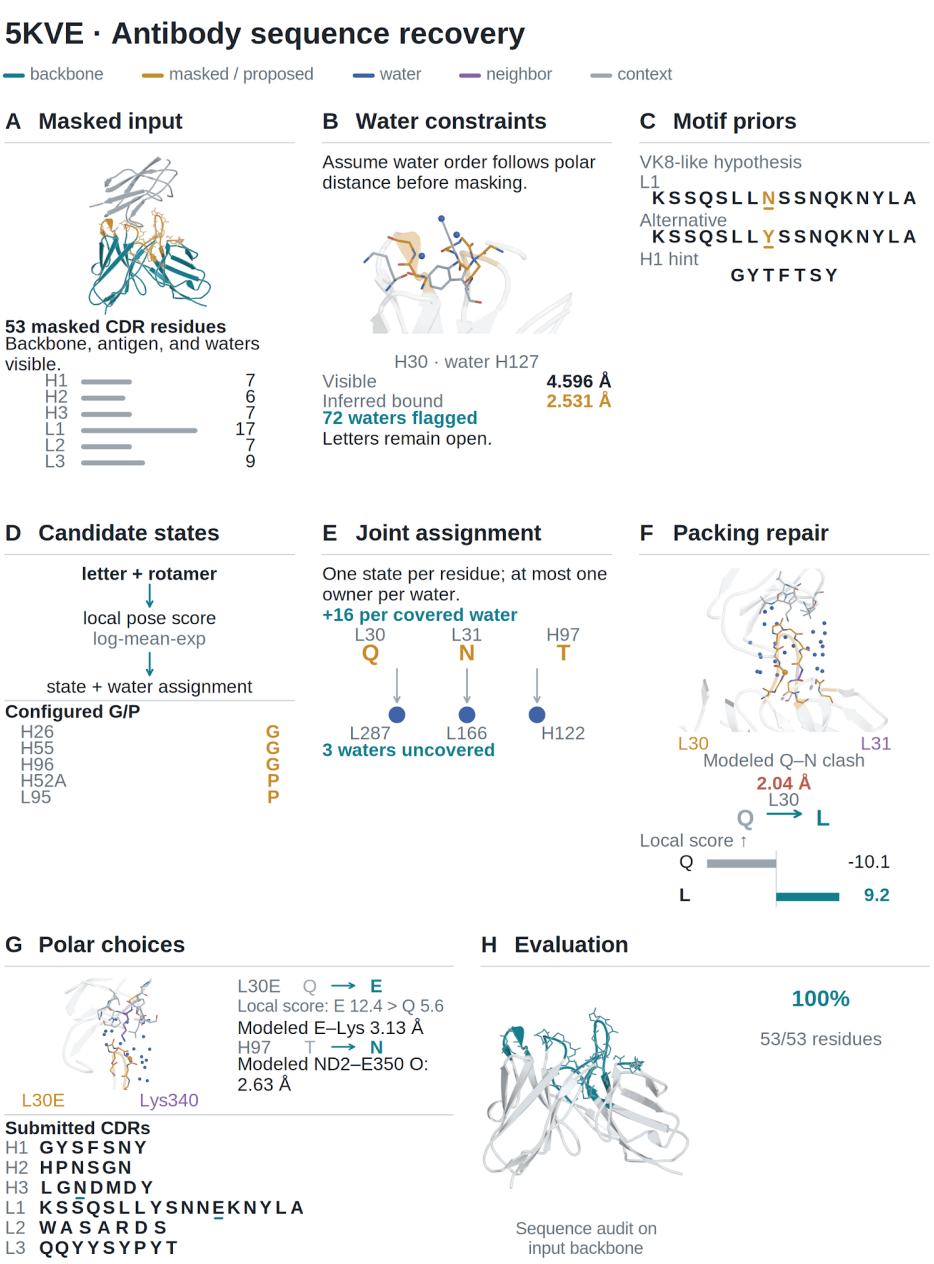


match different

A1: D → native E

After submission, predicted and native sequences are compared. Recovery pools exact matches across all chains.

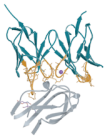
E.2.2 Antibody Inverse Folding



3S35 · Antibody sequence recovery

backbone · masked / proposed · water · neighbor / ion · context

A Masked input and priors



52 masked CDR residues
Backbone, antigen, waters and calcium visible.
H1–3: 7 / 6 / 8 L1–3: 15 / 7 / 9

Motif starting guesses

H1 GYTFTSY
H2 FPGSGS
H3 ????YFDY
L1 RASESVDSYGNSFMH
L2 LASNLES
L3 QQNNEDPLT

H3 retains four open positions.

B Calibrate and sample

Visible-residue calibration

472 visible: 293 fit / 179 withheld

119/179 top-letter matches 66.5%
Internal check on the same structure.

10 replicas: coupled letters + waters
H56: I 61% / S 2%
L91: S 67% / N 1%
Local sampler frequencies (%) ↑

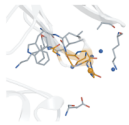
Trial CDRs

H2 FPRSGI
H3 HDGTSFDY
L2 RASNLES
L3 QQSNEDPLT

L1 end FMY → FMH

Trial includes manual sequence judgment.

C Rerun and compare choices



Input H53 region

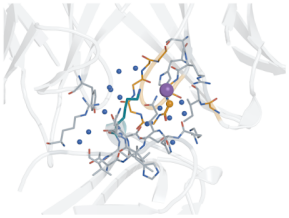
Keep NumPy arrays alive across C pointer call. HDGT, Ile56, Ser91 remained favored.
H53 Arg 44%→42% Gly 5%→3%
Local sampler frequencies: before → rerun

Refit: 124/179 (69.3%)
Later candidate frequencies (%) ↑

H31	D 36%	N 35%
H52	Y 54%	F 46%
H53	G 30%	R 25%
H99	A 50%	S 46%

L50 Arg prior = 0.99999

D Calcium and H3 alternatives



Input calcium / Lys271 pocket

Modeled contacts

Asp96 OD2–Ca H208	2.36 Å
Asp96 OD2–Lys271 NZ	2.65 Å
His95 ND1–H97 N	2.93 Å

H3 alternatives

H3 HDGSSFDY
H3 HDGTAFDY
H3 HDGTSFDY
H3 HDGTGFDY

E Choose Ala and submit

Assumed V(D)J update: Ser/Ala 2.6 → about 2.

Ala99 remained slightly favored.
No single rule fixed H3.

H99
S → A

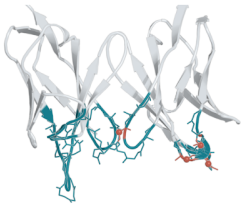
Final heavy-chain revisions

H31	H52	H53
S → D	F → Y	R → G

Submitted CDRs

H1 GYTFTDY
H2 YPGSGI
H3 HDGTAFDY
L1 RASESVDSYGNSFMH
L2 RASNLES
L3 QQSNEDPLT

F Evaluation



Sequence audit on input backbone

90.4%

47/52 residues

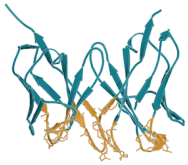
Submitted → native

H28 T → I
H31 D → E
H53 G → E
H55 G → N
H99 A → N

4ZD3 · Antibody sequence recovery

backbone · proposed · water · shared factor · error · context

A Masked input



54 masked CDR residues
Framework and waters visible. No annotated antigen chain.

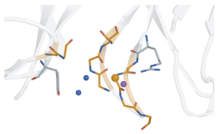
H1	7
H2	6
H3	12
L1	11
L2	7
L3	11

B Motif priors

IGKV1-5 hypothesis
Preferred motifs
H1 **G Y S F T S Y**
L1 **R A S Q S I S S W L A**

Configured prior weights
Heavy preferred 0.955
Light preferred 0.92
H3 open.

C Candidate states



Input · H96 water pocket

17,972 retained states · **267** pair edges
H192 / H196 assigned
H223 shared factor
H96 allowed: H E Q R K Y

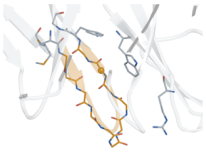
D Neighbor updates

Belief propagation passes compatibility between neighbors.
30 updates · damping 0.45
Conditional weight ↑

Position	Leading	Next
H99	D 0.31	S 0.24
H100A	N 0.39	S 0.16

Mixed H3 rankings.

E L96 reversal



Input · L96 pocket
L96
W → Y
Conditional weight ↑

Prior W / Y	0.65 / 0.075
Initial BP: Y	0.98
Larger BP: Y	0.80
Phase off: Y / W	0.417 / 0.010

F Gene priors

BP/prior × assumed V(D)J
Trimming · insertions · codons
Scale 1 model weight ↑

H99 S	0.718
H100A S	0.446
H100B D	0.309

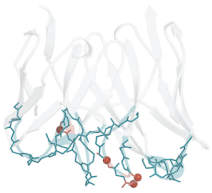
Top sampled H3 prefix
P H Y S S S S D A
754 / 30,000 draws
Prefix lost the lead at scale 0.55.

G Commit candidate

H96 His weight
0.9890–0.9940
across three packing scales

Submitted CDRs
H1 **G Y S F T S Y**
H2 **Y P G D S D**
H3 **P H Y S S S S D A F D I**
L1 **R A S Q S I S S W L A**
L2 **K A S S L E S**
L3 **Q Q Y N S Y S P G Y T**
Fixed letters: packing check.

H Evaluation



Sequence audit on input backbone

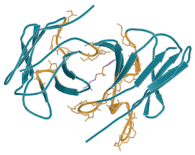
92.6%
50/54 residues

Submitted → native
H98 S → Y
H99 S → D
H100A S → L
L90 Q → H

1DL7 · Antibody sequence recovery

backbone · masked / proposed · water · ligand · context

A Masked input and seed



47 masked CDR residues
Backbone, waters and NCH visible.
H1–3: 7 / 5 / 5
L1–3: 14 / 7 / 9

Template L2 **G T N N R A P**

Motif-based seed
H1 **G F S L T S Y**
H2 **W A G G S**
H3 **X X X D Y**
L1 **R S S T G A V T T S N Y A N**
L2 **G T K H R T P**
L3 **A L W Y S N H W V**
H3 retains three unknown positions.

B Geometry and local scores

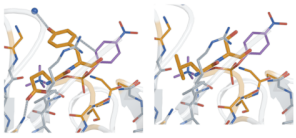
Virtual backbone probes
H53 Cβ–H30 O **2.33 Å**
H97 Cβ–H101 C/O **2.10 / 2.18 Å**
H101 peptide ω **−1.027°**
G53, G97 and P101 proposed

Inferred logSC
H53 −3.1 **H54 +7.1**
+ supports a sidechain; − favors Gly

Local score ↓
H53 **G 0 / D 178.5**
H54 **S −4.8 / A −4.7**
D −4.3 / G 0
H97 **G 0 / A 82.8**
H101 **G 0 / P 1.5**
Asp54: motif-guided, not score-leading.
H101 Gly/cis-Pro open; H53 Ala untested.

C H2 checkpoint and H3 pairs

Separate modeled candidates



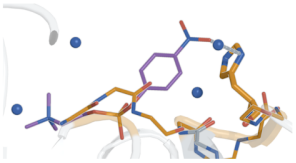
E95/Y96 L95/S96

Joint cost above best ↓
K/S 0
L/S 1.1
E/Y 3.3

E/Y modeled contacts
E → Asn35 2.5 Å
Y → NCH 3.3 Å · water 2.9 Å
Asp54 and E/Y remain open.

H2 checkpoint
H2 **W A G G S** seed
H2 **W G D G S** ↓ A53 → G · G54 → D
H3 **E Y G P Y** working

D Retain the L2 candidate



NCH His53

Modeled His53 on input backbone

Modeled contacts
His ND1–NCH O1N **2.7 Å**
Thr55–H97 O **2.6 Å**
Thr55 Ala/Ser alternatives nearly tied.

His53 versus Lys
Local score ↓
H −17.4
K −16.6

Retained L2
L2 **GTKHRTP**

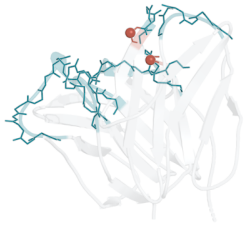
E Priors and submission

H96 weights by assumed gene input
input A **Y 0.38**
restricted D **L 0.52**
H3 **E Y G P Y**
retained
G31: score + motif agreement.
N56: motif choice; S cost 0.1 lower.

Motif-guided revisions
H31 S → G G 0 / S 3.2
H56 S → N N −16.8 / S −16.9
Local score ↓

Submitted CDRs
H1 **G F S L T G Y**
H2 **W G D G N**
H3 **E Y G P Y**
L1 **R S S T G A V T T S N Y A N**
L2 **G T K H R T P**
L3 **A L W Y S N H W V**

F Evaluation



Sequence audit on input backbone

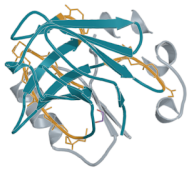
95.7%
45/47 residues

Submitted → native
H56 N → S
H95 E → D

4Z9K · VHH sequence recovery

backbone · proposed · water · neighbor · error · context

A Input and ranking



25 masked residues

H1 / H2 / H3: 7 / 4 / 14

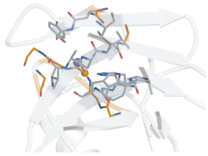
Visible-site held-out matches

231/342 · 67.5%

Same-structure rescoring

CDR letters remain open.

B Cys constraint



Cys50 B97

Configured S–S target

2.036 Å

B97

X → C

Working CDRs · B99 open

H1 **GFTLDDY**
H2 **KDGS**
H3 **RPCYNGVPLIDFGS**

C B99 alternatives

Local score ↑



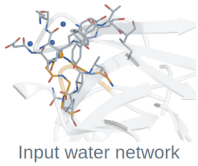
Modeled contacts

Asn OD1 → water B135 2.82 Å

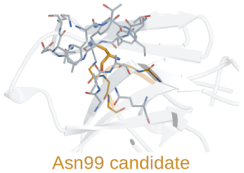
Leu CD1 → Tyr58 ring 3.63–3.66 Å

N / L / R remain open

D Water geometry



Input water network



Asn99 candidate

B135: 4 polar neighbors · 2.68–3.05 Å
OD1–B135 O–B100 N ≈62°

Crowding weakens N/D; Leu remains a judgment.

E Choose Leu

Near-tied scores + water geometry

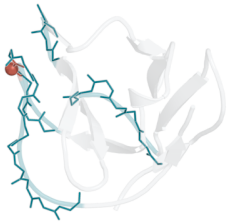
B99

N → L

Submitted CDRs

H1 **GFTLDDY**
H2 **KDGS**
H3 **RPCYLGVPLIDFGS**

F Evaluation



Sequence audit on input backbone

96%

24/25 residues

Submitted → native

B98 Y → F

5E08 · Antibody–RNA sequence recovery

backbone · proposed · water · assumed copy · error · RNA / context

A Masked input



57 masked CDR residues
RNA bases, phosphates, and waters visible.

H1	7
H2	6
H3	17
L1	11
L2	7
L3	9

B Backbone cues

Gly proposals · score ↑

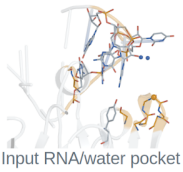
H26	H55	H100
G	**G**	**G**
5.0	**6.0**	**13.2**

L95 · Gly score 11.2

Incoming C–N	1.3443 Å
Filter maximum	1.337 Å
Omega	−0.6°
L95 **P**	cis hypothesis

Outside filter; uncalibrated scores.

C Water assignments



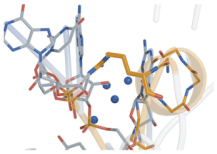
Input RNA/water pocket

Local cost ↓

H53 Y → H223	**−4.8**
H54 Y → H232	**−0.9**
Swapped	27.7 / 44.7

Both Tyr proposals stayed open.
H54 OH–N10 O6: 3.02 Å (modeled)
Assumed crystal context.

D Arg at phosphates



Arg100D candidate
Modeled NH1–RNA OP2

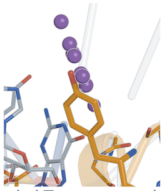
N2	N3
2.76 Å	**2.81 Å**

Local score ↓

R	**-3.7**
S	-1.4
Y	2.6

Fixed background; assumed lattice.

E Self-image check



Recorded Tyr + assumed copy
Corrected indexing check
≈0.8 Å
Minimum separation
Tyr weakened.
Arg copy not tested.

F Motif choices

Local cost ↓

H95 Ser	**−3.6**

Modeled Ser OG–RNA N4 O2/N3
2.79 / 3.04 Å

H100G · lower cost favored R

Background	R	Y
First	**1.8**	4.6
Later	**0.4**	4.1

H95 stored	**S**
H100G stored	**Y**

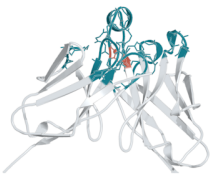
Motif recall; local cost favored R.

G Commit six loops

Submitted CDRs

H1	**GFSISSS**
H2	**YPYYGS**
H3	**SYSSGGSYYRYSYGFDY**
L1	**RASQSVSSAVA**
L2	**SASSLYS**
L3	**QQSYSYPYT**

H Evaluation



Sequence audit on input backbone

96.5%

55/57 residues

Submitted → native

H95 S → Y
H100G Y → R

2J88 · Geometry and aromatic packing

backbone · proposed · water · neighbor / copy · error · context

A Input and motif seed



47 masked CDR residues

Paired antibody with protein A; backbone and surrounding structure were visible.

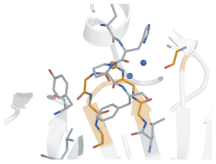
CDR lengths **9/5/7 · 11/7/8**

Framework-recalled seed

H2 **WWDDD**
H3 **YYYAMDY**
L3 **QHHYGTYT**

B Geometry supports Gly

sklearn unavailable → custom tree ensemble. Backbone features ranked residue groups.



Input backbone at H97

Gly enrichment ↑

H26	**8.9**
H31B	**11.5**
H97	**5.2**

Gly proposals: H26 · H31B · H97

C Resample aromatic states

Extra Tyr/Trp rotations improved Trp sampling but retained the Tyr preference at H52.

H52 Trp: −3 → 11 after resampling

Local score ↑

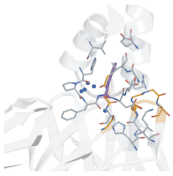


Joint local score ↑

H52Y / L96R	**62**
H52W / L96R	**54**
H52Y / L96S	**62**

H52 Y favored; L96 R/S open

D Choose Tyr at H52



Input H52 pocket

Recalled templates and modeled contacts led to the Tyr choice.

Tyr OH–A145 N, modeled	**3.0 Å**
Trp NE1–A144 NE1, modeled	**2.96 Å**

H2 working revision

H2 **YWDDD**

AMDY remained a motif hypothesis.

E Junction and submission

Junction hypothesis

TGG → GG

C + GG → CGG (Arg)

Restricted Arg / Ser score	**62 / 62**

L96 · seed to submission

Y → **R**

Submitted CDRs

H1 **GFSLSTSGM**
H2 **YWDDD**
H3 **YYGAMDY**
L1 **RASENIYSYLA**
L2 **NAKTLAE**
L3 **QHHYGTRT**

F Evaluation



Sequence audit on input backbone

93.6%

44/47 residues

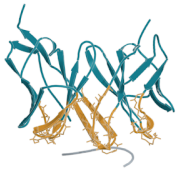
Submitted → native

H98 A → S
H99 M → V
L34 A → T

2QHR · A hypothetical crystal copy reshapes H3

backbone · proposed · water · neighbor / copy · error · context

A Input and motifs



63 masked CDR residues

Paired antibody with peptide P and waters; H3 contains 14 residues.

CDR lengths **7/6/14 · 12/11/13**

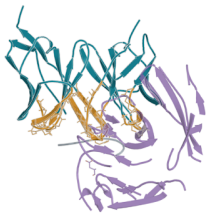
Recalled light-chain motifs

L1 **T L S S Q H S T Y T I E**
L2 **L K K D G S H S T G D**

Identity remained a hypothesis; H3 was still open.

B Hypothetical crystal copy

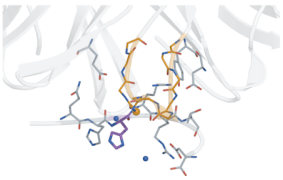
FFT packing supplied a working lattice after cell refinement ended with precision loss.



Working cell: 68.16 × 69.70 × 142.10 Å
Displayed operator

xyz′ = xyz × (1, −1, −1)
+ (34.08, −34.85, 142.10) Å

C Pack a provisional H3



Input H100B · peptide · waters

Motifs and the working crystal context seeded all loops. Tyr100B and Gly100C remained subject to rechecking.

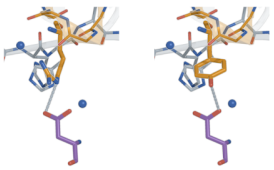
Recorded provisional seed

H2 **S S G G G Y**
H3 **T D Y Y G S S Y G Y A M D Y**

D Revise four H3 positions

Conditional packing supported four replacements.

Local score ↑	
H95 His / Thr	**11.9 / 9.6**
H96 Ile / Asp	**10.7 / 4.3**
H100B Arg / Tyr	**11.9 / −8.8**
H100C Tyr / Gly	**8.8 / 0.0**



Arg candidate **Tyr candidate**

Same hypothetical Asp173 copy

Modeled Arg NE–OD2: 2.84 Å

H3 **H I Y Y G S S R Y Y A M D Y**

E Submit the six CDRs

Local score ↑ · R 13.0 / S 6.0

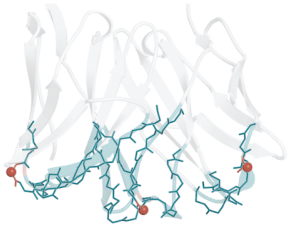
H52A · recorded final choice

S → **R**

Submitted CDRs

H1 **G F S F S S Y**
H2 **S R G G G Y**
H3 **H I Y Y G S S R Y Y A M D Y**
L1 **T L S S Q H S T Y T I E**
L2 **L K K D G S H S T G D**
L3 **G V G D T I K E Q F V Y V**

F Evaluation



Sequence audit on input backbone

95.2%

60/63 residues

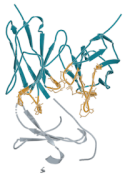
Submitted → native

H28 S → A
H100B R → H
L27 S → R

5T1D · Reopening a Pro restriction

backbone · proposed · water · neighbor / copy · error · context

A Masked input



51 masked CDR residues

Paired antibody with protein B; backbone and surrounding structure were visible.

CDR lengths **7/6/6 · 16/7/9**

Observed H3 positions

H95–H98 · H101 · H102

No residues were added across the numbering gap.

B Constrain H98 to Pro

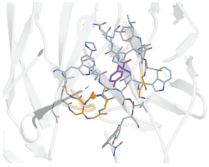
Configured restriction

H98

X → P

H98: Pro only

C Question the Pro fit



Input H98 · Tyr L36

The Pro candidate had a large clash feature and a short modeled approach to Tyr L36.

Clash feature ↑ = worse **45.88**

Modeled Pro CD–Tyr OH

2.27 Å

Restriction still active.

D Remove the restriction

H98: Pro only

↓

ordinary alternatives

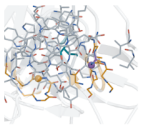
JH4/YAMDY-like motif hypothesis

H3 **G Y A M D Y**

Met98 became a working proposal

E Choose Met and submit

Met was selected from its prior despite Tyr's local lead; Trp at L96 followed the recalled J1 motif.



Input H98/L96 pocket

Conditional score ↑

H98 Y / M **0.0 / −1.5**

L96 H / W **0.0 / −1.5**

L96 W / H · Met favored **0.0 / −2.0**

Submitted CDRs

H1 **G Y T F T D Y**
H2 **Y P S D G S**
H3 **G Y A M D Y**
L1 **R S S Q S L L H S N G N T Y L H**
L2 **K V S N R F S**
L3 **S Q S I H V P W T**

F Evaluation



Sequence audit on input backbone

90.2%

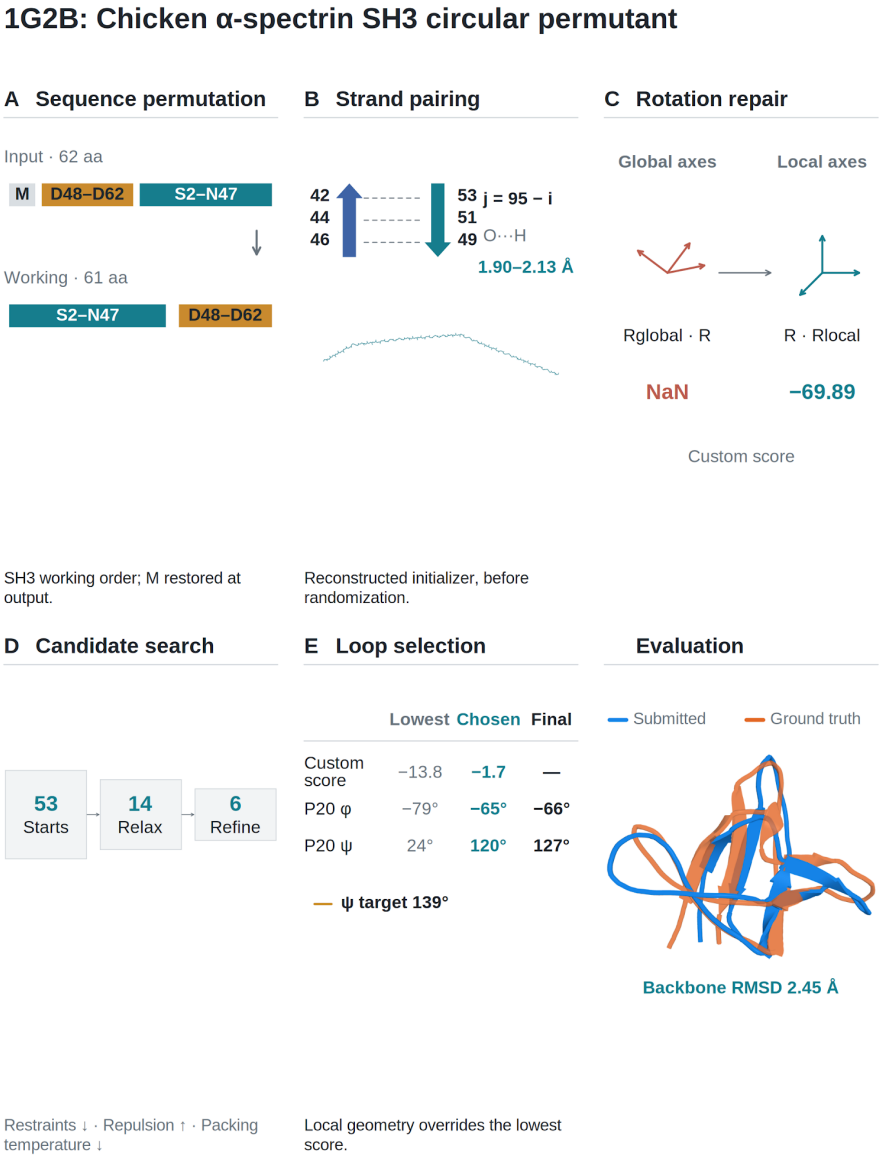
46/51 residues

Submitted → native

H53 S → E
H54 D → S
H56 S → N
H102 Y → F
L96 W → R

E.3 Structure Prediction

E.3.1 Monomer Structure Prediction



3TIP: SasG E–G5₂ biofilm module

A Sequence landmarks

E-like	G5-like
NPETG 33–37	NPLTG 101–105
YGP 50–52	YGP 128–130

17 Pro + 14 Gly = 31 / 132

Repeated motifs support the modeled E–G5 rod.

B Fragment definitions

β-strands

E left

3–9	28–33	39–45

E right

6–14	21–31	42–53

G5 left

53–66	89–101	107–121

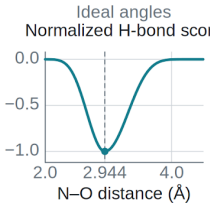
G5 right

64–73	81–91	119–131

Offsets: E (37, 73) · G5 (155, 210)

Four overlapping pieces support the long fold.

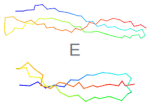
C Fragment search



Seed counts · 38,500

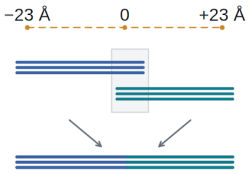
E right	11,000
G5 right	11,000
G5 left	12,000
E left	4,500

Domain Cα: G5



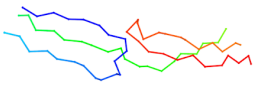
D Fragment assembly

Overlap + 2 axis points



xR + t ≈ y · det(R) = +1

Cα junction



3 retained segments

Axial guides preserve the extended chain.

E Consensus and local repair

Candidate	Custom score	Ensemble Cα RMS
A48	−180.08	1.08 Å
B6 · chosen	**−171.68**	**0.95 Å**

I108 φ/ψ

Before	After
(−116°, −6°) ⟶	**(−88°, −29°)**

Final I108: (−87°, −35°)

Consensus selection precedes 200 local trials.

Evaluation

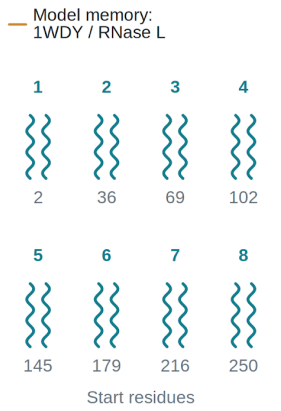
Submitted — Ground truth



Backbone RMSD 3.92 Å

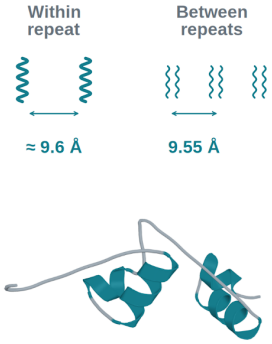
1WDY: Human RNase L ankyrin-repeat regulatory domain

A Ankyrin repeat plan



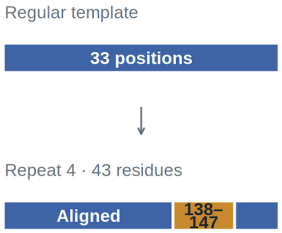
Paired helices form a reusable repeat scaffold.

B Repeat geometry



Reconstructed repeat template; before local fitting.

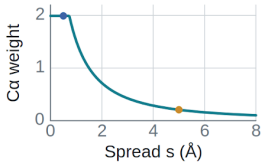
C Insertion mapping



10-residue insertion omitted from the template.

D Consensus weighting

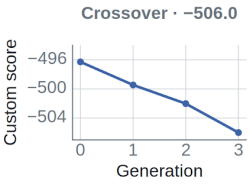
w = clip(q, 0.19, 1.5)^1.7



spread 0.5 Å	spread 5.0 Å
2.0	**0.21**

q = 2.3/(0.8+s) · pool + 5 modeled refs → mean

E Search and loop refinement



Local refinement

93
jobs

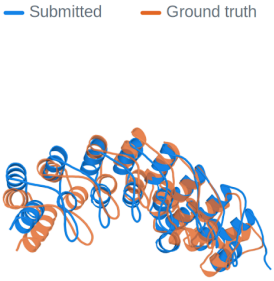
Control −504.8265

Best −504.8425

K146 φ/ψ ≈ (−145°, 19°)

Local refinement uses new restraints.

Evaluation

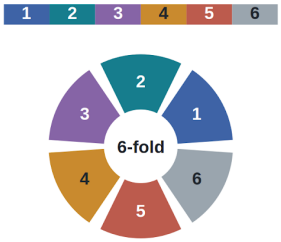


Backbone RMSD 4.60 Å

6l3B: cPizza6-AYW designer beta-propeller

A Sixfold symmetry

252 residues · sequence only · 6 × 42

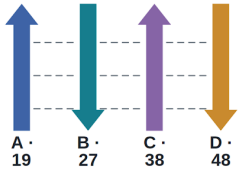


Rz(60° × k)

Assumed cyclic

B Strand pairing

4 strand centers



Pairing

j = a + b + s − i

shift s = −3 … 3 parity e = 0 / 1

7 shifts × 2 parities = 14 registers

C Hairpin screening

A/B registers · 300 starts each

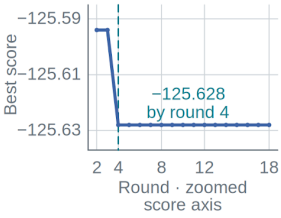
Register 4	**Register 7**
shift −1 · parity 0	shift 0 · parity 1
−16.44	**−14.56**

Custom score

Targets: O···N 2.98 Å · O···H 1.99 Å

D Packing search

24 populations · 18 rounds



Perturb → Relax → Exchange

Reconstructed Cα

Separate searches

Search A



Search B



E Consensus selection

Candidate 0	**Candidate 1**
Lowest	Chosen
−43.8	**−41.6**

Custom score

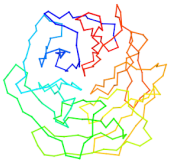
G32 φ = −107°

Shared register (7, 6, 4)

Reconstructed Cα

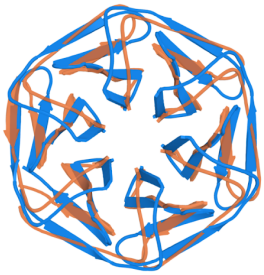
Candidate 1

Before final polishing



Evaluation

Submitted Ground truth



Backbone RMSD 2.93 Å

3QHB: Oxidized Symerythrin

A Repeated metal-binding motifs

179 residues · sequence only

Half 1

E36 … E39 | E70 … H73

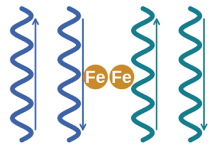
+91 | +91

Half 2

E127 … E130 | E161 … H164

E: glutamate · H: histidine

B Modeled diiron core



Twofold core

Distance targets

Fe–Fe	His N–Fe 2.17 Å
3.84 / 3.45 Å	Glu O–Fe
	2.01–2.22 Å

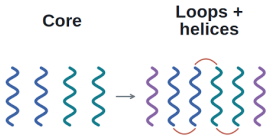
3,200 starts · alternative modeled states

C π-bulge selection

Candidate 1532	Candidate 2767
Unfiltered	π-filtered
Custom score	
−283.6	−253.5
Tyr44 OH···ligand O	
16.1 Å	2.6 Å

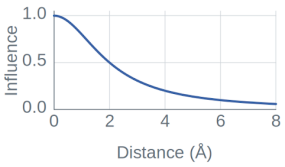
Assumed π H-bonds: 38→43 · 39→44

D Chain assembly



N-terminal	Bridge loops
335,000	500,000

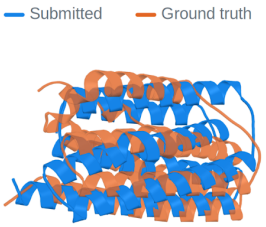
E Consensus refinement



6 updates

w / (4 + d²)

Evaluation



O81–O84: 2.764 → 2.969 Å

Backbone RMSD 5.88 Å

Final scores · 0: −642.1 · 1: −640.9

6NVT: TLA-1 extended-spectrum beta-lactamase

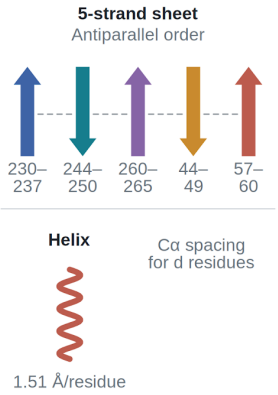
A Conserved motif alignment

276 residues · sequence only

	SxxK · 70	SDN · 130	K[T/S]G · 234
Target	STYK	SDN	KTG
TEM	STFK	SDN	KSG
CTX	STSK	SDN	KTG

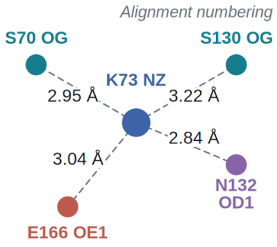
Alignment keys: 70 / 130 / 234

B Fold scaffold



$\sqrt{[(1.51d)^2 + (4.57 \sin 0.866d)^2]}$

C Catalytic pocket



Modeled water

S70 OG · 2.85 Å

E166 OE2 · 2.71 Å

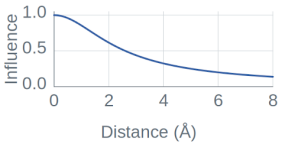
D Joint model search

TEM Modeled ↔ CTX Modeled ↔ Target Modeled

CA/O/CB coupling · 74 Å computational offsets

Sheet register	Target score
A (481,509,309,105)	**−2,405.2**
B (479,509,309,105)	−2,339.7
C (481,510,310,106)	−2,400.4

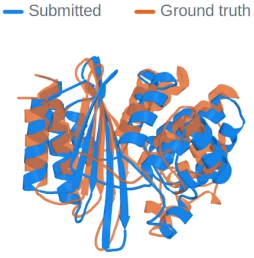
E Flexible consensus



$w = (4 + d^2)^{-0.70}$

3 position updates → 6 distance updates → Relax

Evaluation



Peptide C–N: 1.320–1.341 Å

Backbone RMSD 4.67 Å

4Y0A: Shikimate kinase (AroK)

A Kinase fold

179 aa · no starting coordinates

GPMGAGKT · 15–22

P-loop · Lys21

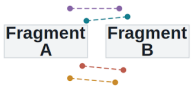
↓

Assumed kinase fold

Memory: H. pylori shikimate kinase

Family recognition narrows the fold search.

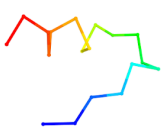
B Fragment closure



Match four virtual points

[1.35x, 1.35y, 1.35xy] Å

x,y = ±1

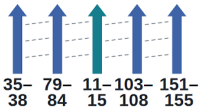


Cα P-loop intermediate

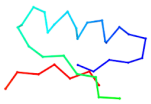
Position and orientation reconnect moved fragments.

C Parallel-strand pairing

Cα target 4.54–5.23 Å



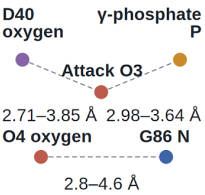
j = i + x − y

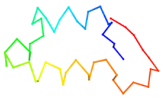


Cα core intermediate

One pairing rule expands three strands to five.

D Modeled pocket



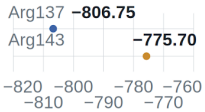


Cα LID · Arg137 candidate

Pocket restraints shape the Arg137 branch.

E Arg143 preference

Before the prior



Custom score · lower is better

+45 for non-Arg143 candidates

A structural prior overrides the Arg137 score lead.

F Candidate consensus



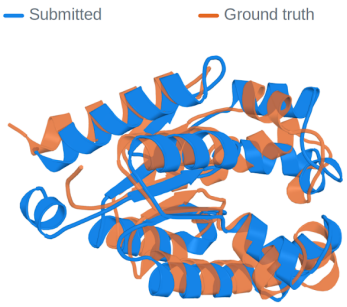
Weight∝ exp(−score / 23)

Stable: stronger · variable: softer

Selected Arg143 · −642.84

Consensus combines with a manual LID shift.

Evaluation



Backbone RMSD 4.66 Å

4ERC: VHZ / DUSP23 phosphatase

A Phosphatase motif

150 aa · no starting coordinates

H	C	A	L	G
94	95	96	97	98

F	G	R	T	G
99	100	101	102	103

HC(X)$_5$R motif

↓

Memory: DUSP23/VHZ

2IMG / 4ERC

Family recognition suggests the initial fold.

B Parallel-sheet geometry

4.76 Å spacing · twist t = 0.121i



16–21 · 90–95 · 36–41 · 57–62

N–O hydrogen bond: 2.97 Å



Cα loop · before sheet assembly

An idealized sheet anchors the catalytic loop.

C Modeled catalytic pocket

Asp65 OD1 — Ligand O: **2.97 Å**

Arg101 NH2 — Phosphate P: **3.88 Å**

Arg101 CZ — Glu44 CD: **4.07 Å**

Assumed contacts orient the acid loop and arginine.

D Candidate search

Mutate φ / ψ

↓

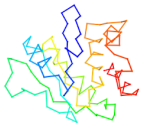
Recombine Swap region

↓

Relax Pack + minimize

Crossover 0.46

48	1,250
12 parallel	
Configured runs	Trials per run



Rounded Cα · candidate i14

Fragment reuse explores larger structural changes.

E Consensus register

f · 24.5%

z · 40.0%

i · 35.5%

92% robust estimate + 8% weighted mean

Weight = 0.168 / (1 + σ$^{1.6}$); σ = spread

↓

(28,21, 41,60)	(28,21, 40,60)
−343.5	−320.8

Consensus favors the lower custom score.

F Terminal assembly

N end → Core ← C end

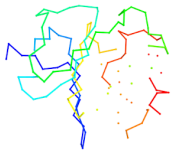
↓

Chain 1–150

96 terminal candidates

Model join RMS

2.357 → 0.077 Å

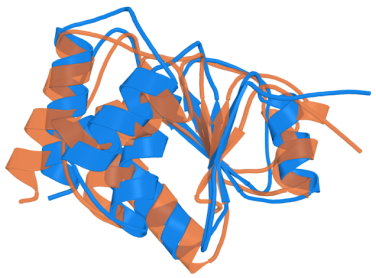


Rounded Cα · 126/150 positions

Terminal search follows register selection.

Evaluation

Submitted — Ground truth

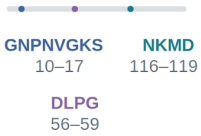


Backbone RMSD 3.60 Å

2WJH: FeoB GTP-binding domain

A FeoB fold

166 aa · no starting coordinates



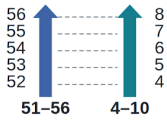
Assumed FeoB fold

Memory: 2WJG apo · 2WJH GDP

Motifs and model memory narrow the starting fold.

B P-loop scaffold

Register example: −48



Cα target	4.85 Å
Lys16–phosphate	2.98 Å
Ser17–Mg	2.12 Å

Assumed contacts align the P-loop and strands.

C Modeled guanine contacts

3.04 Å
Asn116 — Guanine N7

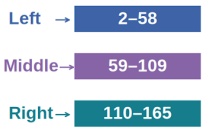
2.89 / 2.92 Å
Asp119 carboxylate — Guanine N1 / N2

2.945 Å
Residue 146 N — Guanine O6

Residue 146 contact: G5 option

Later contacts connect NKMD/G5 to the pocket.

D Regional assembly



Crossover spans · 2–60 cuts before 27, 38, 50



Cα regional candidate



Rounded pre-final assembly

Regional solutions assemble into a larger candidate.

E Candidate search

12
Runs

562
Trials per run



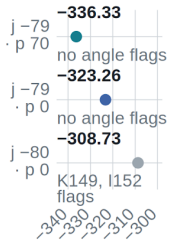
45% crossover

64% torsion changes

Recombination retains useful regions while exploring.

F Connected-chain ranking

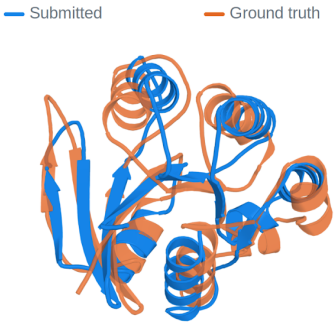
Connected chain · 1–166



Custom score · lower is better

Angle checks complement score ranking.

Evaluation



Backbone RMSD 4.44 Å

Full Monomer Structure Prediction Gallery

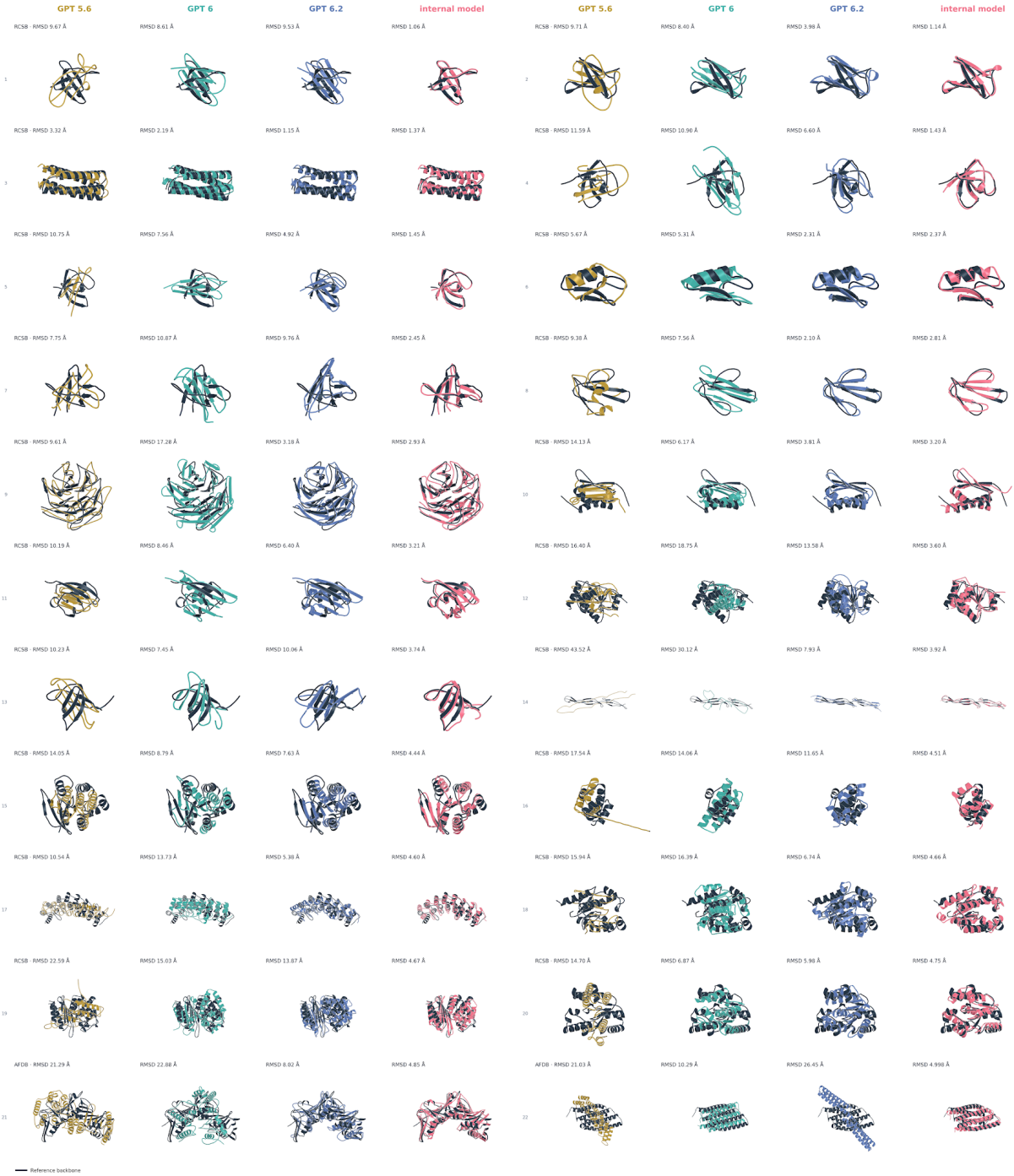
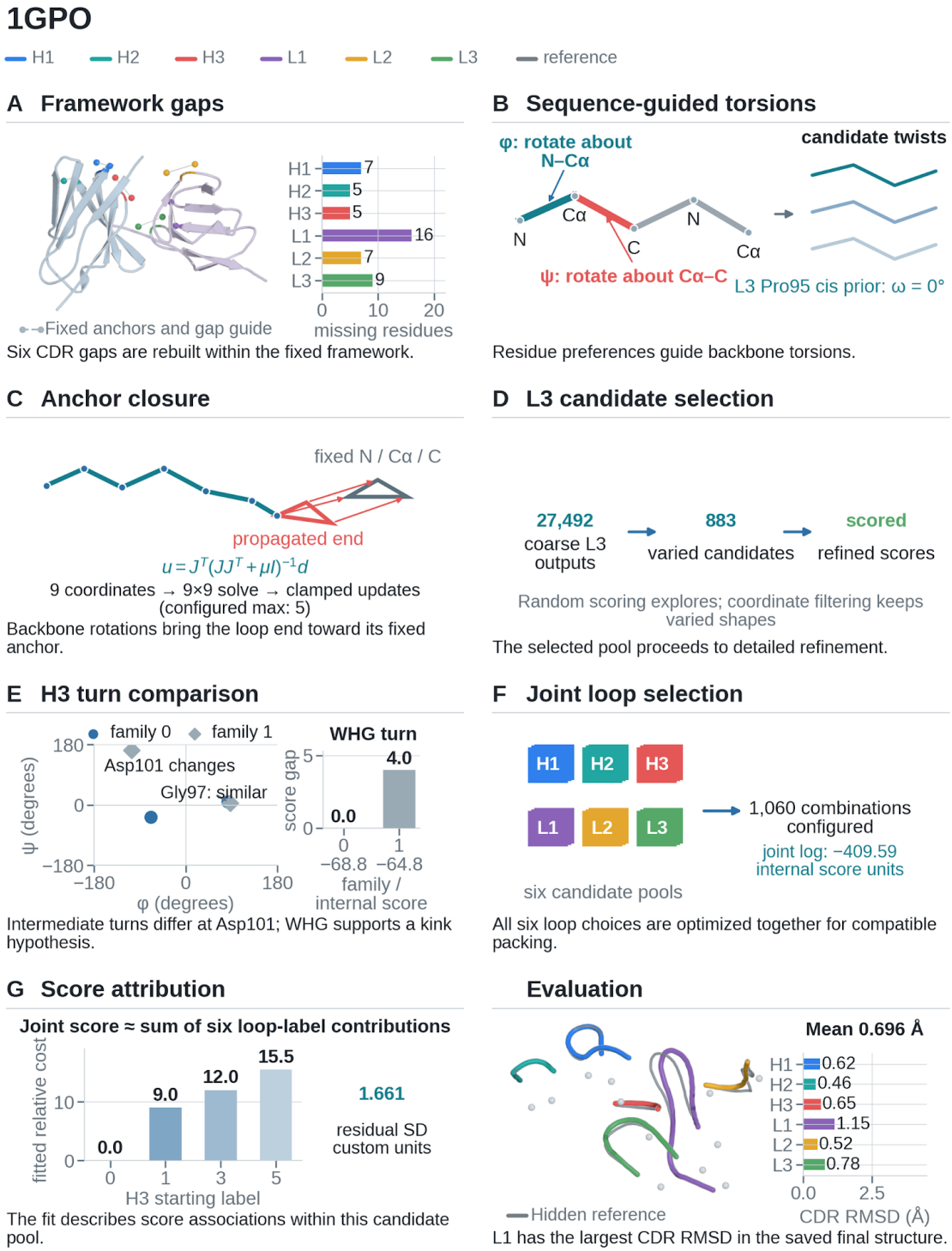


Figure E1. Four-model monomer structure predictions for all 22 selected cases, extending the examples in Figure 22H. Colored predictions overlay dark reference backbones; labels report RMSD (Å) after global N/CA/C/O backbone alignment. These selected illustrative cases use model-specific budgets.

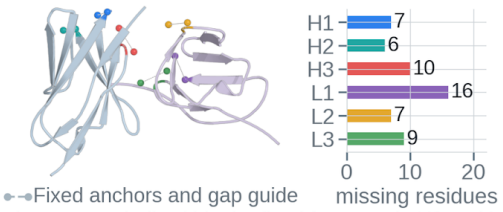
E.3.2 CDR Structure Prediction



1RUQ

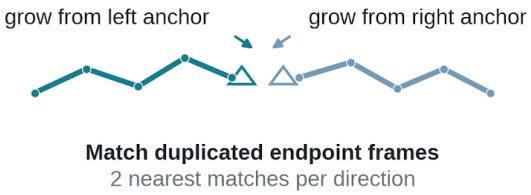
H1 H2 H3 L1 L2 L3 reference

A Framework gaps



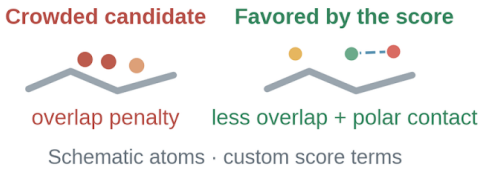
Six loops are rebuilt within the fixed framework; H3 and L1 form the largest gaps.

B Growth from both anchors



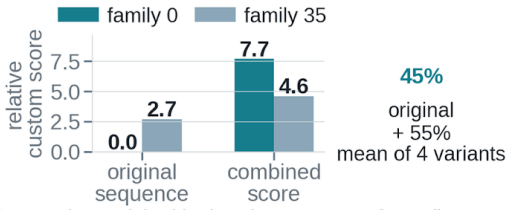
Loop halves grow from both anchors; matching junction positions and orientations identifies compatible joins.

C Atomic environment scoring



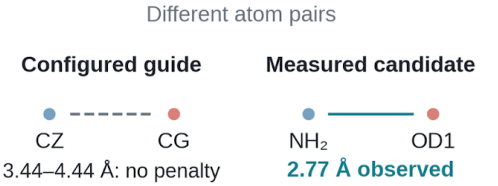
Temporary side chains support scoring of packing, burial, charge, and hydrogen bond geometry before backbone refinement.

D Sequence variant challenge



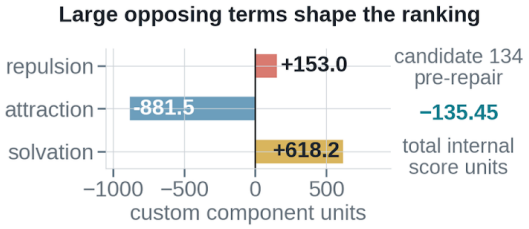
Four variants picked by hand test scores after refinement. Homology is unverified; displacement is from each starting candidate.

E H3 contact prior



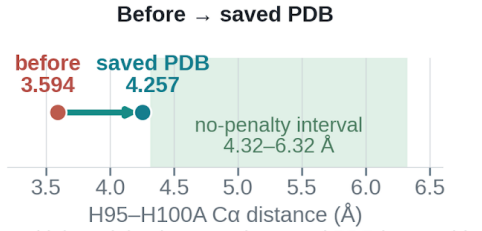
An Arg94–Asp101 guide is imposed; a different atom pair is measured in an auxiliary candidate with side chains.

F Score balance



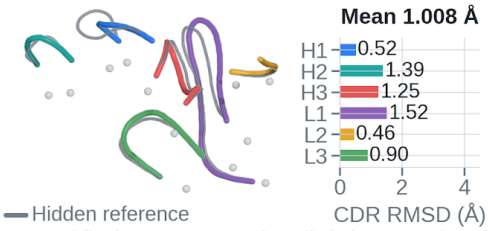
Candidate 134's printed components require stage weights to interpret the total.

G H3 crowding repair



After multiple weight changes, the saved PDB has a wider Cα gap, still below target. The final write call is absent.

Evaluation

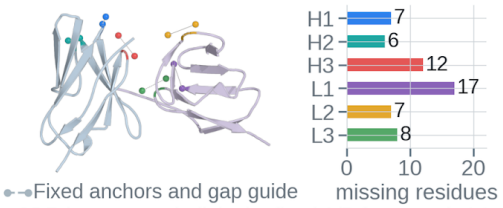


The saved final structure contains all six loops; L1 has the largest CDR RMSD.

1IL1

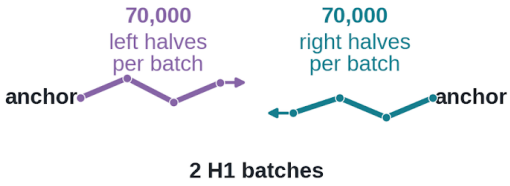
H1 H2 H3 L1 L2 L3 reference

A Framework gaps



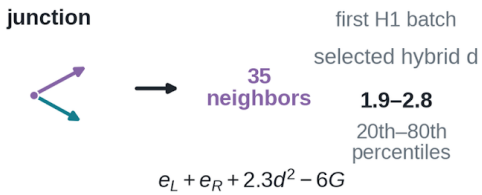
Six CDR loops are rebuilt in the fixed A/B framework; L1 and H3 form the largest gaps.

B Growth from both anchors



Shorter loop halves grow from both anchors, reducing construction to matching compatible junctions.

C Junction matching



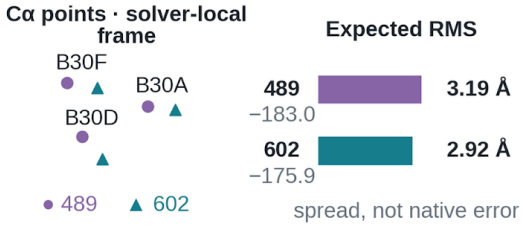
A hybrid distance matches position and orientation; Gumbel score noise retains alternative joins.

D Packing and backbone relaxation



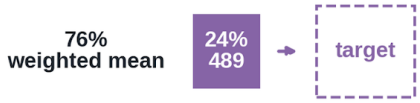
Temporary side chains are packed, then the backbone is relaxed to compare candidates in a fuller atomic environment.

E Score and candidate agreement



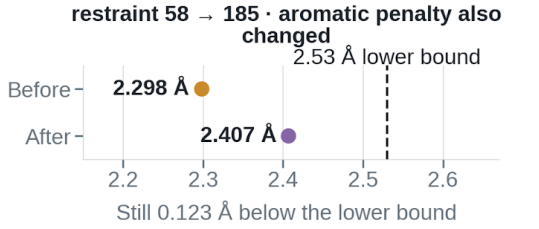
Candidate 489 has the lower internal score; 602 has the smaller RMS spread among candidates.

F Compromise target



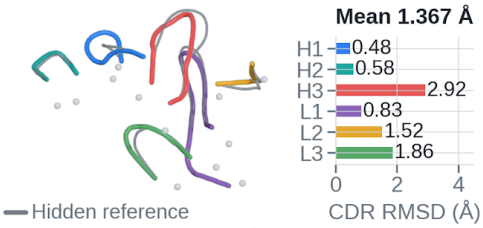
The L1 target blends weighted mean coordinates with candidate 489; downstream optimization remains configured only.

G Trp50 crowding repair



The rerun increases Cβ···H distance, still below target; the constructed hydrogen is absent from the final PDB.

Evaluation

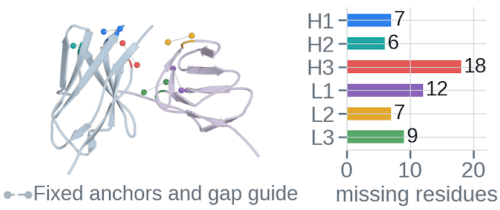


Mean CDR RMSD is below 2 Å; H3 has the largest loop error in the saved final structure.

1HZH

H1 H2 H3 L1 L2 L3 reference

A Framework gaps



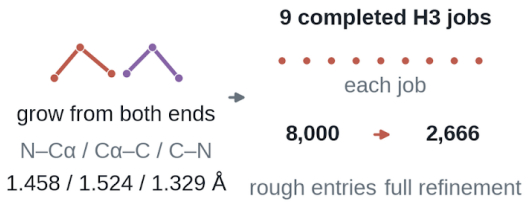
Six CDR loops are rebuilt in the fixed framework; H3 is the longest gap, with no antigen or ligand guidance.

B H3 exposure hypothesis



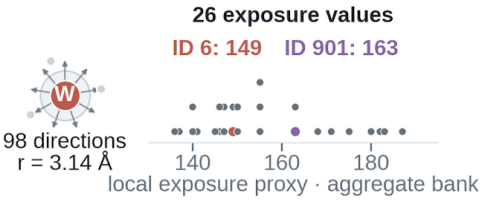
A b12 analogy favors an exposed Trp100 tip; antibody identity and bound state are unconfirmed; no ligand was supplied.

C H3 candidate screening



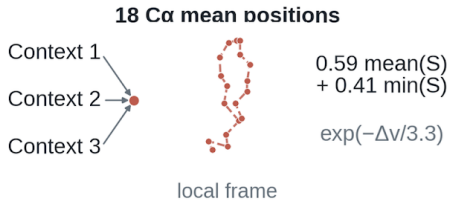
Nine H3 jobs filter processed entries by shape diversity; counts do not establish unique structures across jobs.

D Local Trp100 exposure



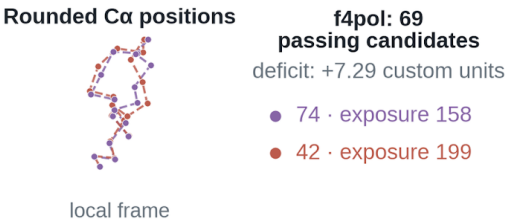
Counts of free directions around Trp100 use only H3 neighbors within 12 Å as a local exposure proxy for ranking.

E Coordinate guide from three contexts



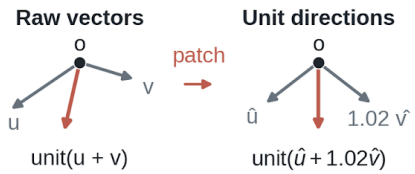
Mean H3 Cα positions form a coordinate guide across three contexts, using exposure, peptide angle, kink, and sampling preferences.

F Filtered H3 candidates



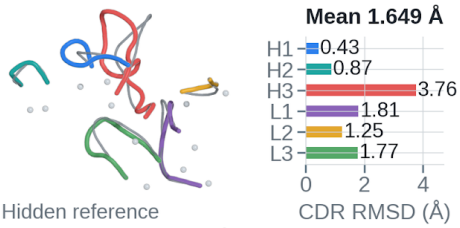
Candidate 42 has higher exposure and is configured as a seed; its link to the final backbone remains unverified.

G Carbonyl direction correction



Carbonyl bond vectors and derivatives are updated for later refinement; a full rerun of the earlier search is unverified.

Evaluation

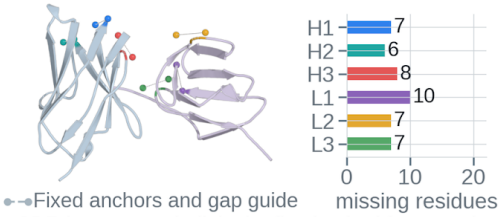


Mean CDR RMSD is below 2 Å; H3 has the largest error. Candidate 42's link to the final backbone is unverified.

1MIM

H1 H2 H3 L1 L2 L3 reference

A Framework gaps



Six CDR loops are rebuilt in the fixed paired framework; H3 and L1 require different shapes.

B Loop-half matching



mean position + two directions

9-value junction feature

Junction position and orientation match compatible loop halves before refinement.

C H3 candidate selection

209,658
batch 1 matches

206,473
batch 2 matches

subsample + refine

100,000
coarse outputs

Pair matches are counted before subsampling

Two batches produce junction matches; subsampling and refinement form the coarse H3 pool.

D Water accessibility



86
probe directions

2.95 / 3.24 Å
radial positions

Geometric probes estimate polar-atom access to water for candidate scoring.

E Polar satisfaction

best H-bond
extra contacts
water access

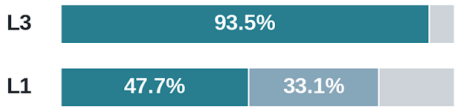
deficit + saturation

3.64
penalty SD score units

0.15
correlation with prior score

Buried unsatisfied atoms and excess contact credit are penalized to rerank candidates.

F Candidate density correction



normalized model weights · remaining groups in gray

Density correction reduces repeated-shape influence; L3 concentrates, while L1 remains split.

G Backbone checks

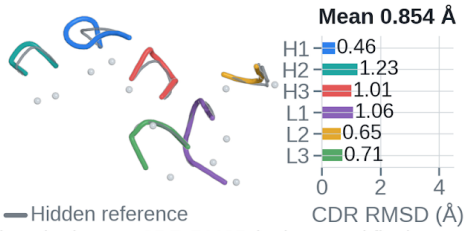


+32.4°
H3 base torsion

1.328–1.330 Å
C–N link range

After writing, backbone checks find short reciprocal H96/H99 contacts and nearly uniform C–N links.

Evaluation

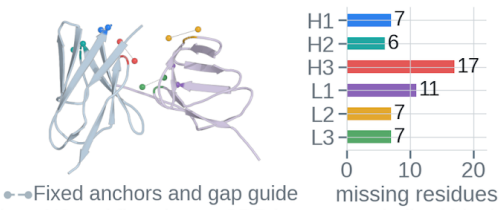


H2 has the largest CDR RMSD in the saved final structure.

1DFB

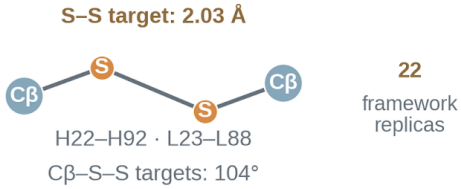
H1 H2 H3 L1 L2 L3 reference

A Long H3 gap



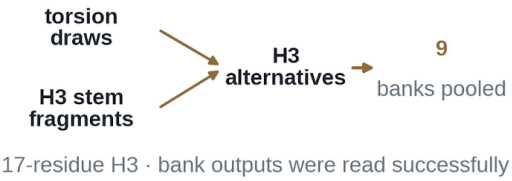
A long H3 and five shorter CDR loops are rebuilt within the fixed framework.

B Disulfide initialization



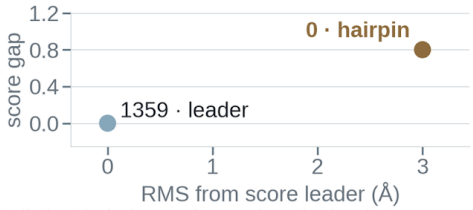
Targets initialize two auxiliary disulfides before packing; they are not final measured distances.

C H3 candidate generation



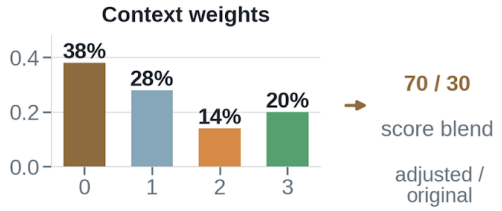
Candidate banks from torsion-sampled loops and transplanted H3 stems are pooled for comparison.

D H3 shape comparison



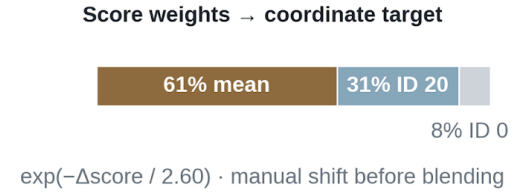
A distinct hairpin nearly matches the leader's score; the model favors its proposed SGG turn.

E Context-weighted ranking



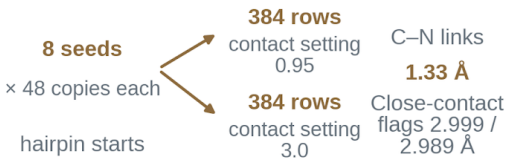
Four short-loop contexts test H3 candidates; a chosen adjusted/original score blend ranks them.

F Guided coordinate target



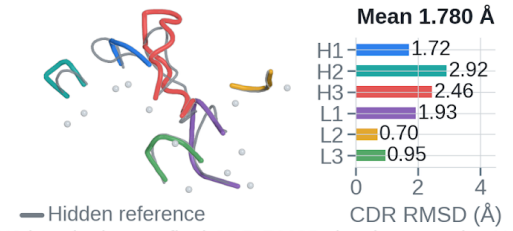
Score-weighted coordinates are manually shifted and blended with two candidates for guided refinement.

G Hairpin refinement



Two targeted fits refine chosen hairpin seeds under H3 guides; checks still flag close contacts.

Evaluation

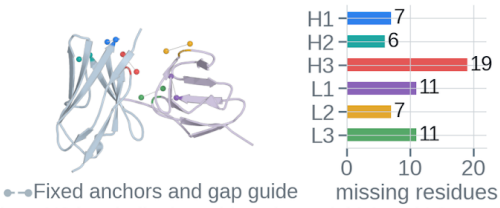


H2 has the largest final CDR RMSD despite extensive H3 search and guided refinement.

1RZ8

H1 H2 H3 L1 L2 L3 reference

A Longest H3 gap



A long H3 is rebuilt alongside substantial L1 and L3 gaps in the paired framework.

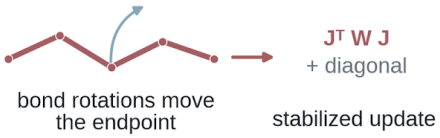
B Sequence-guided torsions

residue angle wells ▸ random backbone twists ▸ varied starting shapes

raise base well weights to power 0.75 → apply optional residue bias normalize for sampling · φ/ψ noise scale 0.76

Residue-dependent torsion mixtures and occasional cis prolines diversify starting conformations.

C Anchor closure



Backbone rotations place atoms; endpoint-sensitive curvature guides stabilized updates, while closure penalties pull toward anchors.

D Contact and packing score



7.5 Å pair cutoff

The custom score balances local packing, directional H-bonds, torsion preferences, and endpoint closure.

E Kink-weighted target

Combine shape classes

38% extended | 62% kink

weighted coordinate target

Code defines a kink-weighted coordinate target; execution and linkage to the later mean file remain unverified.

F Candidate ranking

Combined ranking

score gap in custom units

candidate	399	537
score gap	**0.0**	**+11.0**
subtotal	−293.8	−300.4
unresolved part	24.3	42.7
base torsion	+55°	−171°

The selected candidate wins the combined ranking despite a worse subtotal; one component remains unidentified.

G Guided geometry refinement

Selected candidate 399

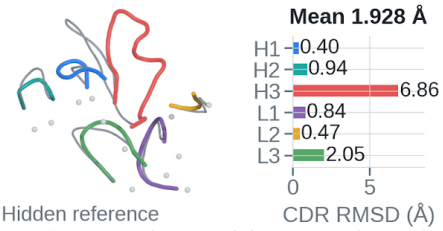
310 endpoint weight ▸ 1,410 endpoint weight ▸ 10,400 endpoint weight

Torsion guides and local distance penalties

Temporary side-chain check: 2.499 Å contact remains

Torsion guides and tighter closure refine the selection; a temporary side-chain check still finds a close local contact.

Evaluation

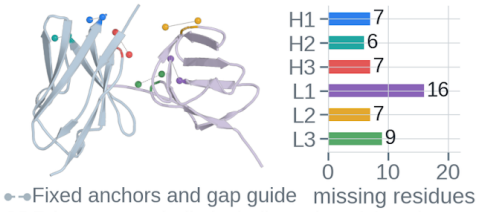


Mean CDR RMSD masks a much larger error in H3, the longest loop.

2C1O

H1 H2 H3 L1 L2 L3 reference

A Framework gaps



Six CDR loops are rebuilt, including a long L1 and a shorter H3.

B Backbone candidate generation

residue angle mixtures ▸ combine angle blocks ▸ two-stage refinement

correlated phi/psi draws · fixed framework anchors

Residue-dependent angle blocks are combined; shortlisted backbones are refined with geometric scores.

C Atom-pair score diagnostic

custom model → Atom-pair diagnostic

atom i · atom j
repulsion · attraction
H-bonds · electrostatics

Inspect which contacts contribute to a candidate score

Four atom-pair score components expose which contacts influence candidate selection.

D Water accessibility



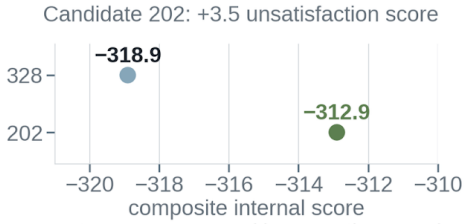
150
probe directions

2.99 Å
radial positions

0.88 × polar deficit enters reranking

Geometric water probes reduce penalties for accessible polar atoms; buried unsatisfaction receives a squared deficit penalty.

E L1 candidate comparison



Contact patterns are compared between the score leader and a lower-ranked L1 alternative before selection.

F Asx–Gly motif hypothesis

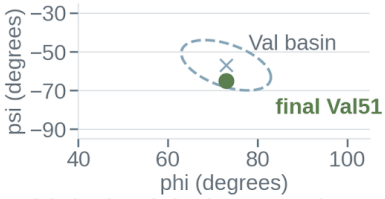
Asx turn + Gly bulge

Y	S	N	G	K
30A	30B	30C	30D	30E

Candidate 202 chosen · Gly30D φ ≈ +90°

The model favors an Asx-turn/Gly-bulge interpretation despite a worse internal score, using Gly30D's positive φ as a motif cue.

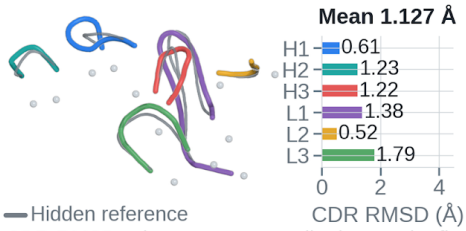
G Val-prior refinement



73° / −65°
phi / psi

A global Val angle basin supports the L2 turn hypothesis; final Val51 geometry remains prior-conditioned.

Evaluation

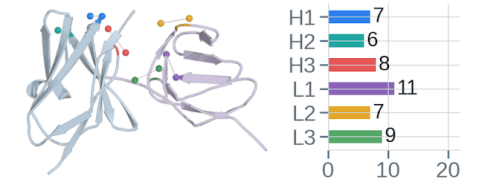


Per-CDR RMSDs show uneven contributions to the final mean.

1CIC

H1 H2 H3 L1 L2 L3 reference

A Framework gaps



Six CDR backbones are rebuilt within the fixed paired framework.

B H3 candidate selection



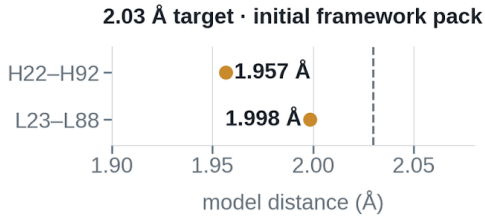
Sequence priors and closure filters narrow the H3 pool for refinement.

C Side-chain packing



Metropolis cooling explores rotamers broadly, then favors lower scores before continuous relaxation.

D Framework disulfides



Initial disulfide distances differ from their targets and do not describe the backbone-only submission.

E Val29 anchor selection

Minimum atom-pair distances: Val29 CB/CG1/CG2 to selected target side-chain atoms

Candidate	Phe71 ring atoms	light-chain residue 2	Variant score
2	9.6 Å	7.7 Å	0.0
42	5.0 Å	4.6 Å	+2.9

variant score in internal units

Candidate 42 brings Val29 closer to the scaffold and gains target weight despite candidate 2's better variant score.

F Pro–Gly turn selection

Residue	φ	ψ	
Pro52A	−65°	+133°	2.86 Å contact
Gly53	+97°	−24°	alternative: +1.7 internal score units

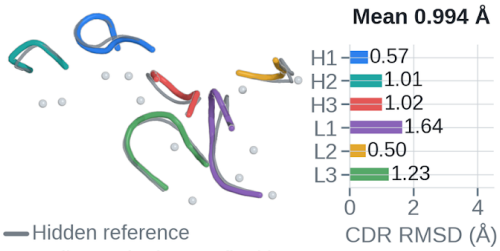
Restricted refinement yields Pro–Gly torsions suggesting a type-II turn; another joint candidate remains competitive.

G Weighted target refinement



Weighted loop targets guide combined optimization, followed by restricted refinement and backbone writing.

Evaluation

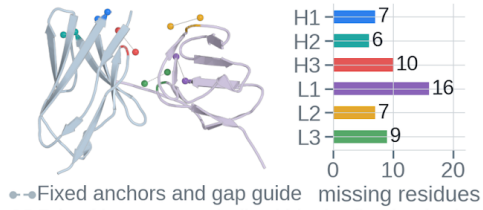


L1 contributes the largest final loop error.

1IGI

H1 H2 H3 L1 L2 L3 reference

A Framework gaps



Six CDR backbones are rebuilt within the fixed paired framework.

B Ligand hypothesis

anti-digoxin 26-10? → **Trp100 presentation**

unconfirmed identity — no ligand supplied

An anti-digoxin analogy guides Trp100 presentation.

C Guided H3 search

distance-guide upper bounds

12 / 15 / 12 Å → **12,760** returned candidates

toward H35 / H50 / L91

Trp100 distance guides focus the returned H3 candidate pool.

D Dummy-point refinement

10 dummy points → **4 dummy points**

initial proxy — narrower proxy

Trp100 model slot 10 → dummy point 1
configured distance window: 4–6.7 Å

Dummy-point windows guide cavity refinement as imposed targets, not measured ligand contacts.

E H3 base comparison

Candidate	Internal score	Base dihedral	Contact proxy
Z3	**−71.05**	**15.2°**	**3.7 Å**
Z536	**−69.26**	**−179.7°**	**6.9 Å**

another candidate scores −71.32

Under dummy-point scoring, Z3's proposed kink and an extended alternative score similarly; another candidate scores slightly better.

F Joint refinement

Z3 → **4 cycles**

selected H3 — all six loops

dummy-point proxy active throughout

The selected H3 guides joint backbone refinement.

G Backbone connectivity

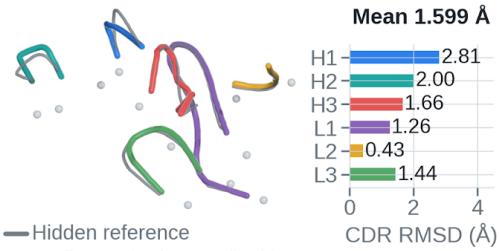
1.326–1.342 Å



61 printed C–N junction distances

The model checks printed junction distances to assess backbone connectivity after refinement.

Evaluation

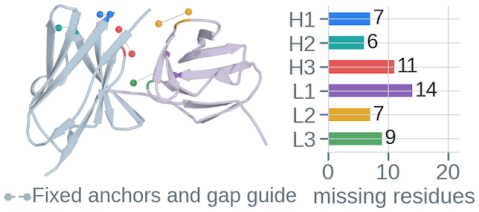


H1 contributes the largest final loop error.

1PG7

H1 H2 H3 L1 L2 L3 reference

A Framework gaps



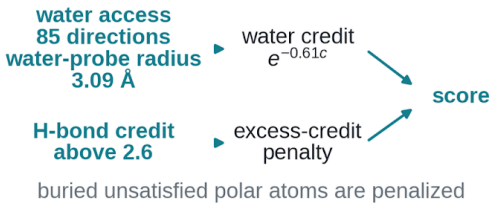
Six CDR backbones are rebuilt within the fixed paired framework.

B Loop candidate selection



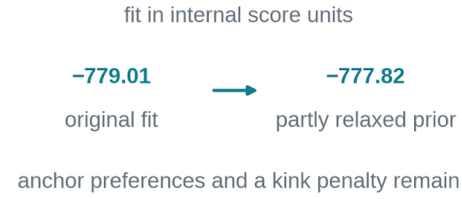
Sequence priors, coarse optimization, and diversity filtering narrow the pool for refinement.

C Polar satisfaction



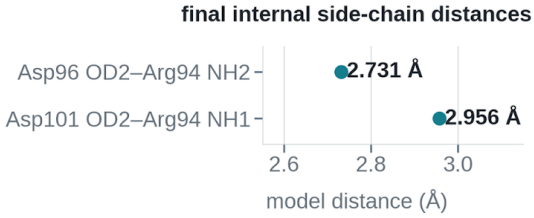
A geometric water-access proxy and excess hydrogen-bond penalties refine candidate ranking.

D H3 prior sensitivity



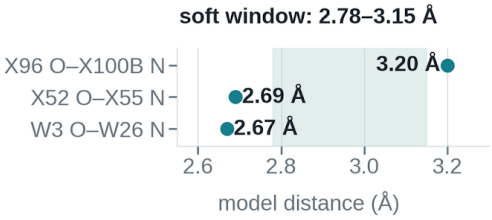
A comparable H3 motif persists under a partly relaxed shape prior.

E Auxiliary charged contacts



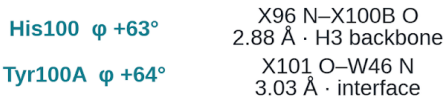
Arg94 lies near Asp96 and Asp101 in the auxiliary side-chain model, not the backbone-only submission.

F Soft contact restraints



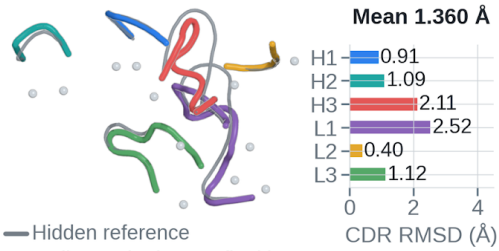
All three O–N measurements lie outside the soft target window.

G H3 backbone strain



His100 and Tyr100A retain unusual positive φ angles alongside the specified backbone N–O contacts.

Evaluation

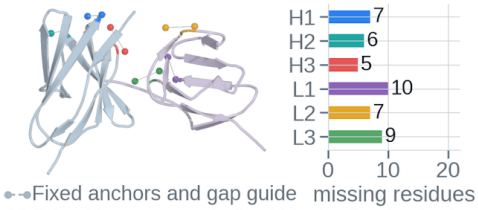


L1 contributes the largest final loop error.

1PSK

H1 H2 H3 L1 L2 L3 reference

A Framework gaps



Six CDR backbones are rebuilt within the fixed paired framework.

B Short H3 sampling

38,000 configured H3 sampling iterations → **28,278** written H3 candidate rows

Sequence priors and two closure checks select H3 candidates.

C Compact H3 selection

Residue	φ in refined pool	
Lys95	**+57°**	**+4.1 units**
Ser96	**+62°**	flipped alternative
Phe97	**+70°**	

Three positive φ angles characterize the compact H3; the flipped alternative has a worse internal score.

D L3 sequence sensitivity

L3 sequence	Preferred motif gap
QQRSGYPFT	**0.0**
QQYNSYPYT	**+0.5**
QQYYSYPWT	**+0.5**

gap in internal score units · cis-Pro held fixed

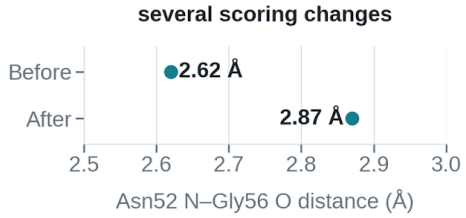
Sequence tests give mixed support for the competitive L3 motif.

E Joint loop selection



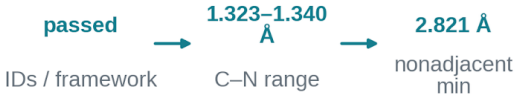
Selected loop families are combined; the leading joint candidate carries compact H3 into refinement.

F Contact repair



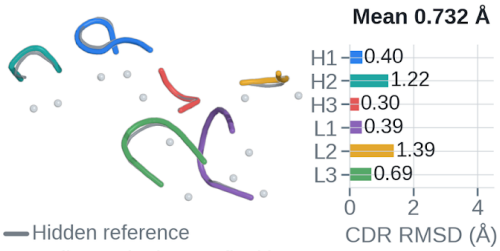
Local reoptimization increases separation after changes including backbone/Cβ repulsion.

G Backbone checks



Readback checks residue identity, framework preservation, junction lengths, and nonadjacent distances.

Evaluation



L2 contributes the largest final loop error.

E.3.3 Protein Docking

2WYT

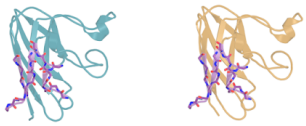
A B

A Read backbones



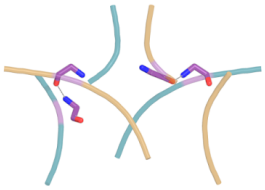
Two independent backbones contained only N, Cα, C, and O. Rigid placement preserves each scaffold while missing side chains are constructed.

B Choose interface



Under a SOD1 hypothesis, purple marks Phe50, Gly114, and the terminal strand. This patch and approximate twofold symmetry restrict the docking search.

C Revise strand pairing



Objective ↓

First **215.27**

Revised **16.33**

Revised Ile151–Gly114 / Gly51–Ile151 pairing gave N–O contacts near 3 Å. Scores use the same form with changed restraints; the backbone view is approximate.

D Build local chemistry

2.032 Å
Disulfide target

2.055 Å
His–metal target*

109.5°
Zn angle target

No ions
Protein output

Cu–His63 used a looser restraint.

Rotamers were packed under disulfide and metal site restraints. These are configured targets; ions were used only as modeling context.

E Compare packing choices

B Ile113

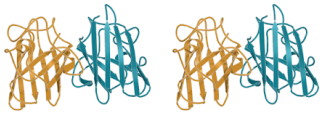
χ_2

Run 8	Run 2
-69°	172°
−1566.3	−1566.0

Among 15 saved ensembles, nearly tied runs gave different B Ile113 angles. Run 8 was chosen, and tighter contacts were favored after inspection.

Evaluation part

Submitted | Ground truth



Panel C uses rounded pose parameters; panel E is an angle schematic.

1BGS

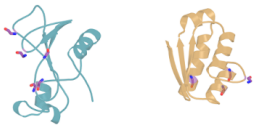
A B

A Read scaffolds



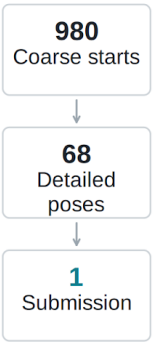
Inputs with only backbone atoms define scaffolds of 110 and 89 residues in separate frames. They must be placed together while preserving both shapes.

B Map charged partners



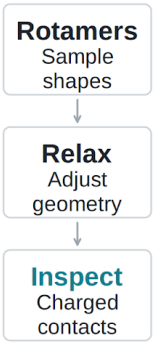
Purple marks proposed charged patches. Remembered barnase–barstar contacts pair basic Arg/Lys with acidic Glu/Asp to focus which surfaces should meet.

C Search rigid poses



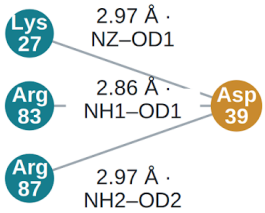
Rigid rotations and translations were scored for surface complementarity and overlap. The coarse job completed, followed by detailed packing and ranking of retained poses.

D Pack charged contacts



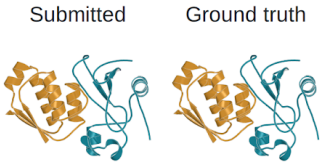
Rotamers sample side chain shapes, and continuous relaxation seeks compatible local geometry. Charged contacts are then inspected in this trace workflow.

E Inspect contact network



Intermediate Asp39 contacts link three basic partners. A close competitor differs by only 0.10 Å RMS in partner Cα positions, showing local agreement.

Evaluation part



Panels C–E are trace diagrams.

4Y0Y

A B

A Read backbones



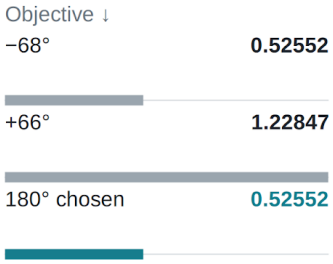
The larger and smaller chains arrived in unrelated frames with backbone atoms only. Docking must determine their association and reconstruct missing side chains.

B Propose catalytic geometry



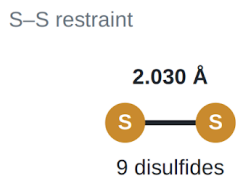
Under a trypsin–inhibitor hypothesis, the model aimed Lys15 toward an acidic pocket and its carbonyl oxygen toward the oxyanion hole, which stabilizes negative charge.

C Test lysine states



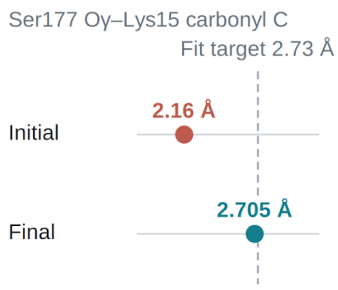
Rigid fitting tested three Lys15 states, with two nearly tied scores. The selected pose put Ser177 Oγ 2.16 Å from Lys15 carbonyl C versus a 2.73 Å target.

D Build covalent links



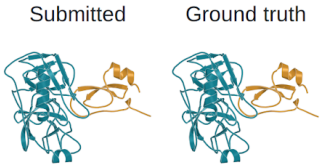
Side chains were built while nine intended cysteine pairs were restrained toward disulfide geometry. The sulfur distance shown is a configured target.

E Refine and validate



Refinement enlarged the catalytic gap. The validator inferred connectivity without moving atoms and passed nine bond checks; a radius screen still found 0.515 Å overlap.

Evaluation part



Panels C–E are trace diagrams.

2DWS

A B C

A Align three copies



Three backbones share a sequence of 328 residues but start in independent frames. Aligning corresponding atoms enables comparison before assembling all three chains.

B Find an axis clue

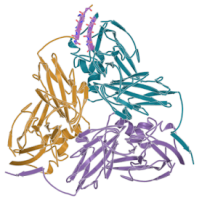
Phase coherence ↑

Peak **0.515**

Next **0.322**

After aligning copies, the model tested how projected coordinates matched a repeating grid. A 0.001 Å coherence peak suggested an axis, with uncertain origin.

C Close the trimer



Docking score ↑

Preferred **425.34**

Next peak **119.15**

A fixed 120° turn reduced docking to two shifts. Four highlighted contacts in the approximate closed trimer proxy match the trace at 0.01 Å precision.

D Pack interface atoms

2.05 Å
His–Cu target

1.24475
Returned fit cost

Rotamers
Side-chain prior

No metal
Protein output

Rotamer packing combined steric, hydrogen bond, charge, and virtual copper terms. The metal distance is a target; the cost is a returned fit result.

E Polish and check

Objective ↓

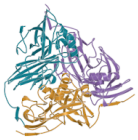
−467.98 → **−499.32**

140 iterations completed

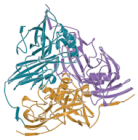
Optimization lowered its custom objective. Final checks found a substantial interface and backbone error below 0.001 Å, with warnings about some proline angles remaining.

Evaluation part

Submitted



Ground truth

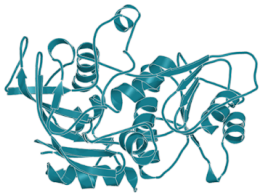


Panel C repeats backbone A, so original B/C differences are absent; panel E shows endpoints only.

3FPL

A B C D

A Inspect four backbones

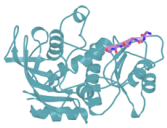


4 inputs

A B C D

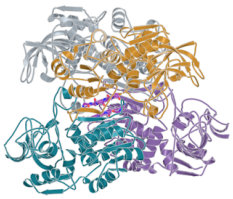
Four backbones share a 351-residue sequence but arrive in independent frames. Representative A is shown; all copies require docking and side-chain completion.

B Pair exposed strands



An alcohol dehydrogenase fold hypothesis guided pairing of exposed strands 287–292. A twofold rotation restricted the dimer search.

C Repair and assemble



Min backbone/pseudo-Cβ (Å)

Saved **1.536**

Recovered **2.664**

A diagnostic exposed overlap in a wrongly saved pose. Fixing the save condition recovered the dimer, then D2 docking assembled the tetramer.

D Pack side chains

2.127 Å
His59
NE2–virtual Zn

2.277 Å
Cys37
SG–virtual Zn

Rotamers and custom contact scores packed side chains around virtual zinc and ligand models. The two distances are internal measurements, not restraint targets.

E Check written geometry

<0.001 Å
max backbone deviation

1,237
A–B atom pairs <5 Å

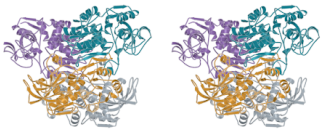
543 / 400
A–C / A–D atom pairs <5 Å

2.879 Å
Gly292 C–Gly284 O

Readback checked backbone preservation and interface size. Extensive contacts remained alongside a tight Gly292 C–Gly284 O contact.

Evaluation part

Submitted | Ground truth

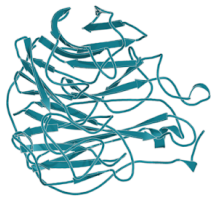


The D2 proxy repeats the aligned A/B/C/D mean; letters mark generated positions, and virtual sites remain model assumptions.

4H53

A B C D

A Inspect four backbones



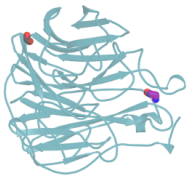
4 inputs

A B C D

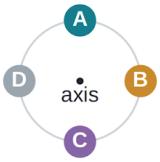
Four backbones share a 388-residue sequence but arrive in independent frames. Representative A is shown; all copies require docking and side-chain completion.

B Propose a C4 ring

Coral N terminus · Purple Asp32/D113

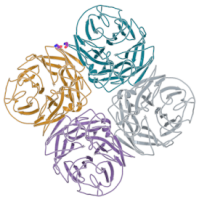


90°



Under a neuraminidase hypothesis, the N terminus and file Asp32 guided radial placement of the aligned A/B/C mean in a C4 ring.

C Compare ring placements



Objective ↓

Best -180.2

Next -90.2

Clash and contact scores ranked C4 placements. The proxy repeats the aligned A/B/C mean; file A Thr374 N–B Asn119 O is unusually short at 2.465 Å.

D Vary packing environments

8 ensemble products

0 / 3 / 4 ligand settings

Shared ion-proxy builder

Graph rotamer search

Glycans, ligand counts and energy weights varied across designs; one ion-proxy builder was shared. Graph search packed rotamers in these assumed environments.

E Select by consensus

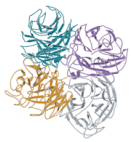
Pool entering round

1	42479
2	29468
3	29691
4	29688

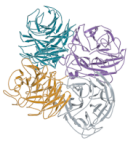
The first bar counts the initial pool; later bars count rebuilt pools. Checks then reported 0 heavy-atom and 45 approximate-hydrogen warnings above 0.4 Å overlap.

Evaluation part

Submitted



Ground truth

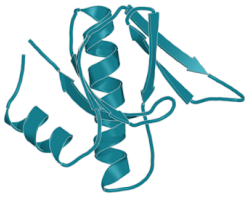


Proxy letters mark generated positions, not final assignments; packing environments are assumed.

4FP5

A B C D E

A Inspect five backbones

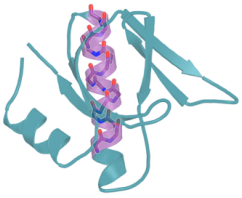


5 inputs

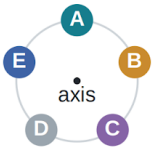
A B C D E

Five backbones share a 98-residue sequence but arrive in independent frames. Representative A is shown; docking must place all copies and complete side chains.

B Propose a C5 ring

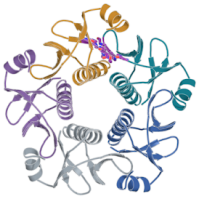


72°



A toxin B pentamer hypothesis used helix 60–73 to define a shared ring frame. Fivefold symmetry restricted how neighboring copies could meet.

C Close the ring



Objective ↓

Best -96.751

Next -49.339

A cyclic search favored compact packing. Logged proximity checks use a virtual point beyond N, so they are not physical hydrogen-bond lengths.

D Pack and link cysteines

Cys10–Cys81 target

2.031 Å

S S

5 disulfides

Rotamers rebuilt side chains while a Cys10–Cys81 covalent constraint targeted 2.031 Å sulfur separation. This configured target is not a measured bond length.

E Compare copy order

All-pairs ↓

AEDCB 19.53

ABCDE 23.30

Sheet-only ↓

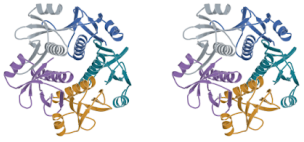
ADBCE 1.46

AEDCB 1.84

Small rigid adjustments compared similar but nonidentical backbones. The all-pairs and sheet-only objectives favored different orders, leaving assignment uncertain.

Evaluation part

Submitted Ground truth

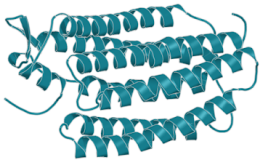


The C5 proxy repeats aligned D; letters mark generated positions, and the disulfide diagram shows a configured target.

6REX

A B C D E

A Inspect five backbones



5 inputs

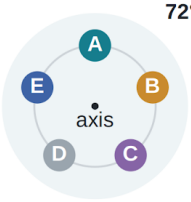
A B C D E

Five backbones share a 273-residue sequence but arrive in independent frames. Representative A is shown; docking must place all copies and complete side chains.

B Define a membrane frame



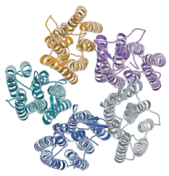
72°



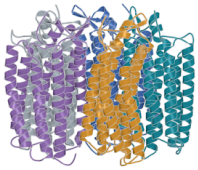
Ends of seven proposed membrane helices defined a rigid frame for A; NDQ positions are highlighted. C5 constrained axis position and tilt; KR2 identity remained a hypothesis.

C Dock the C5 ring

Top



Side



Overlap and packing scores selected a C5 pose. Its approximate reconstruction keeps all five backbones rigid; logged proximity matches within rounding, and ABCDE is provisional.

D Pack assumed chemistry

2.81 Å
Tyr–His

2.37 Å
Asp–virtual Na

Rotamers and custom steric/polar scores packed side chains around virtual sodium and retinal sites. The displayed contacts are intermediate and conditional on those assumptions.

E Resolve serine crowding

Custom score comparison
Ser70 χ_1

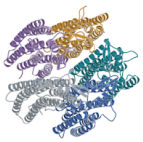


Lower score	Selected
−746.87	−746.08

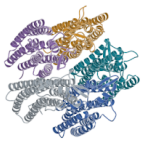
The Ser70 state with lower custom score placed its oxygen near two others. Inspection favored trans with a 2.74 Å Asp contact; KR2 Ser70 = PDB Ser68.

Evaluation part

Submitted



Ground truth

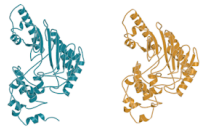


Proxy ABCDE is provisional; membrane and chemical sites are schematic, and rotamers are intermediate.

1U6R

A B

A Inspect two backbones



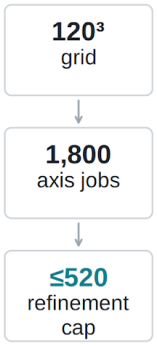
Two copies share a 380-residue sequence but arrive in unrelated frames with only backbone atoms. Docking must determine their association before side chains are built.

B Propose a C2 dimer



A creatine kinase dimer hypothesis imposed a half turn around a shared axis, with axial shift limited to 3.65 Å. The protein identity remained unconfirmed.

C Scan translations by FFT



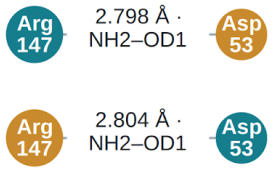
FFT compares gridded shapes across translations, then clusters and refines survivors. The corrected run reports 1,800 axis jobs; ≤520 is the configured refinement cap.

D Correct the overlapping pose



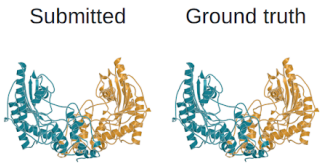
An initial 1.26 Å collision involved side-chain beads absent from these backbone views. Revised scoring changed the interface; reconstruction matches 515 logged proxy contacts.

E Inspect polar contacts



After packing and refinement, final trace checks measured reciprocal Arg147–Asp53 contacts. The diagram preserves the atom pairs and their measured distances.

Evaluation part

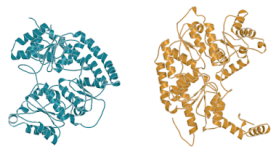


Reconstructed backbone proxies are approximate and omit the modeled side-chain beads used in contact checks.

6OP4

A B C D

A Choose AB

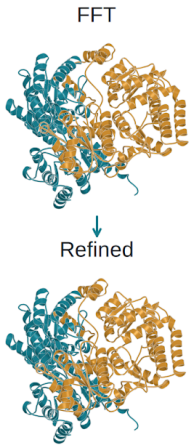


FFT score ↑

AB	**2094.8**
AC	1824.5
BD	1706.8

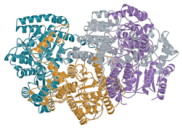
FFT scores shape fit and hydrophobic contact. Representative A/B inputs are shown; AB ranks highest.

B Refine AB



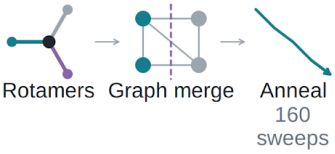
Rigid optimization increases overlap weight 7 → 42 to reduce crowding while preserving each backbone. Matched views show the pose change.

C Copy AB



Dock a second AB dimer through a 177° turn. This proxy creates C/D from A/B, so differences in the original C/D backbones are absent.

D Pack side chains



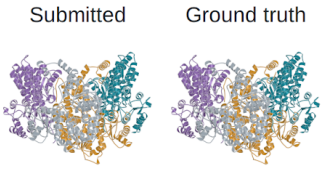
Rotamers sample side-chain shapes; a graph cut combines compatible choices. Annealing searches for better packing under steric and polar scores.

E Repair Tyr–Cys



An added minimum-distance restraint separates oxygen and sulfur to reduce local crowding. The points show before/after trace measurements.

Evaluation part



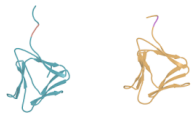
Backbone docking stages are reconstructed; FFT grids differ.

3R0E

A B

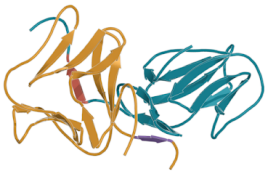
A Propose strand exchange

Coral A100–103 · Purple B105–107



The QxDxNxVxY motif and GNA lectin identity were hypotheses used to propose reciprocal strand exchange. Seven donor proxy and carbonyl pairs guided the fit.

B Choose the register



Backbone pairs <2 Å

95,100	0
97,100	3
95,102	5
93,98	16

Of 16 rigid fits, only (95,100) had no backbone pair below 2 Å; its unrounded minimum was 2.448 Å. Bars show four fits; the view is reconstructed.

C Restrain sheet contacts

7
N–O restraints

2.85 Å
target

8Σ(r−2.85)²
restraint cost

Steric + polar
custom terms

Seven N–O targets constrain the sheet interface while custom steric, solvation and polar terms rank side chains. The 2.85 Å target is configured, not measured.

D Explore packing states

Packing ensemble

210 → **420**

seeds → endpoints

Configured: 80k + 185k + 3×107k proposals

Monte Carlo varied score weights and six pose variables (σ=0.092), producing 420 endpoints from 210 seeds. Completion does not establish equilibrium or convergence.

E Choose Lys102

42%
cluster 0 χ3 179°; score 0.0

21%
cluster 1 · selected χ3 66°; score 1.2

48 / 0.69
cap / damping

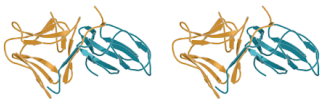
0.000507
earlier residual

Earlier frequencies, χ3 only, and relative custom scores

Min-sum messages repack neighbors; with consensus gain 1.77 and manual bonus 3.8, inference selected Lys102 cluster 1. Shown scores precede those bonuses.

Evaluation part

Submitted | Ground truth



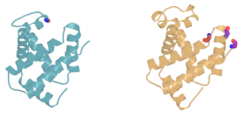
The register view is reconstructed; the Lys102 comparison uses earlier trace states.

1A0U

A B C D

A Read globin priors

Coral Tyr42/Asp99 · Purple His146/Asp94



Representative A/B inputs show two chain types; C/D share their respective sequences but have distinct backbones. Highlighted contacts encode a T-like globin hypothesis.

B Recover grid rotations

0.001 Å grid

0.2916 → **5.9e−12**

A grid residual

After renaming lat.so to liblat.so and updating the loader, a twofold-guided Fourier retry recovered A's rotation. The grid score measures numerical alignment.

C Rank pair roles

C–B opposite · score ↓

#14 **-3.775**

#2 **102.458**

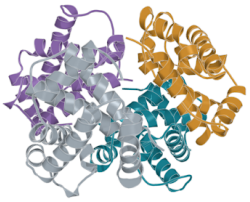
C–D dimer · score ↓

#4 **-7.122**

#11 **41.603**

Across 24 axis-permutation rotations per role, optimized shifts favored C–B opposite and C–D dimer. Compare custom scores only within each role; swaps remained plausible.

D Fit chain shifts



5 pair targets

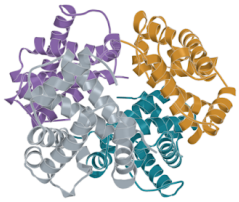
4 chain shifts

LSQ joint fit

2.01 Å backbone minimum

Least squares reconciled AB, CD, AC, AD and CB shifts into four chain translations. Rounded rotations and shifts make this an approximate backbone reconstruction.

E Inspect His146



3.129 Å NE2–Asp94 before

2.932 Å NE2–Asp94 biased

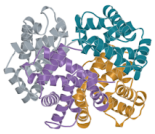
2.82 Å residue 144 O · lost

+101° His146 χ2

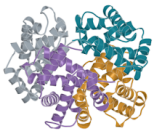
Later v2 shifts reconstruct the backbone view. In a separate biased His146 trial, NE2–Asp94 shortened to 2.932 Å, but the 2.82 Å contact to residue 144 O was lost.

Evaluation part

Submitted



Ground truth



Intermediate molecular views use rounded trace parameters; His146 values come from a separate biased trial.

1BGS

A B

A Mark charged patches



Under a barnase–barstar hypothesis, Arg59–Glu76 and Lys27/Arg83/Arg87–Asp39 define proposed charged patches. Purple marks them on separated inputs.

B Optimize rigid pose

3
translations

φ = q/8
3 rotation variables

2.65wr²
overlap penalty

6 Å
neighbor cutoff

r = min(gap, 0)

L-BFGS-B adjusts three translations and three scaled rotations while keeping both backbones rigid. A 6 Å KD-tree search supports custom restraint, overlap and polar terms.

C Search diverse starts

9
workers

↓

980
starts

↓

68
poses

A parallel search completed 980 starts and retained 68 diverse poses for detailed ranking. Job completion does not establish convergence of every optimizer.

D Pack and relax

Rotamers
discrete choices

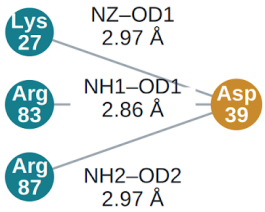
Relax
continuous geometry

−458.17
leader total

−50.68
leader cross

Rotamer sampling and continuous relaxation refined side-chain packing. Candidates were ranked by total + 0.22 × cross; lower custom scores were preferred.

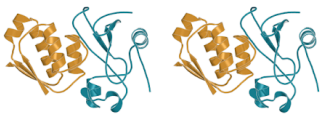
E Inspect Asp39 contacts



Measured Asp39 contacts support the leading detailed pose; Arg59–Glu76 is 2.90–2.96 Å. The next pose differs by 0.10 Å in partner Cα positions, showing only local agreement.

Evaluation part

Submitted Ground truth

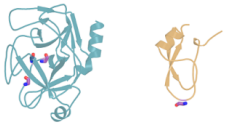


Packing and contact graphics summarize recorded intermediate scores and measurements.

4Y0Y

A B

A Propose catalytic contacts



Under a trypsin–BPTI hypothesis, B Lys15 aimed its side chain toward A Asp171 and its carbonyl O toward the charge-stabilizing oxyanion hole. The image shows separated inputs.

B Compare Lys15 orientations

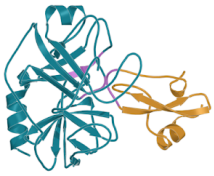
Fit objective ↓

−68°	0.52552
+66°	1.22847
180°	0.52552

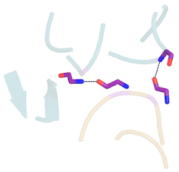
Three Lys15 orientations were scored by a custom rigid-fit objective. −68° and 180° were effectively tied at 0.52552, so selecting 180° was not a strong score preference.

C Inspect initial fit

Initial pose



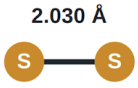
Initial interface



Printed parameters reconstruct the initial backbones, matching seven N–O distances within 0.005 Å. Modeled Ser177 Oγ–Lys15 carbonyl C was 2.16 Å versus a 2.73 Å target.

D Build disulfide links

Disulfide target

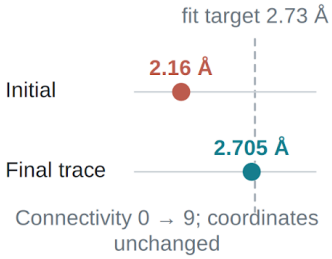


9 disulfides

Rotamers rebuilt side chains and torsions relaxed under steric and polar terms. Nine intended disulfides were restrained toward 2.030 Å; this is a configured target.

E Repair bond topology

Ser177 Oγ–Lys15 C



MDTraj read zero disulfides; inferred connectivity let the unchanged nine-bond check rerun. Refinement still left a 0.515 Å radius overlap.

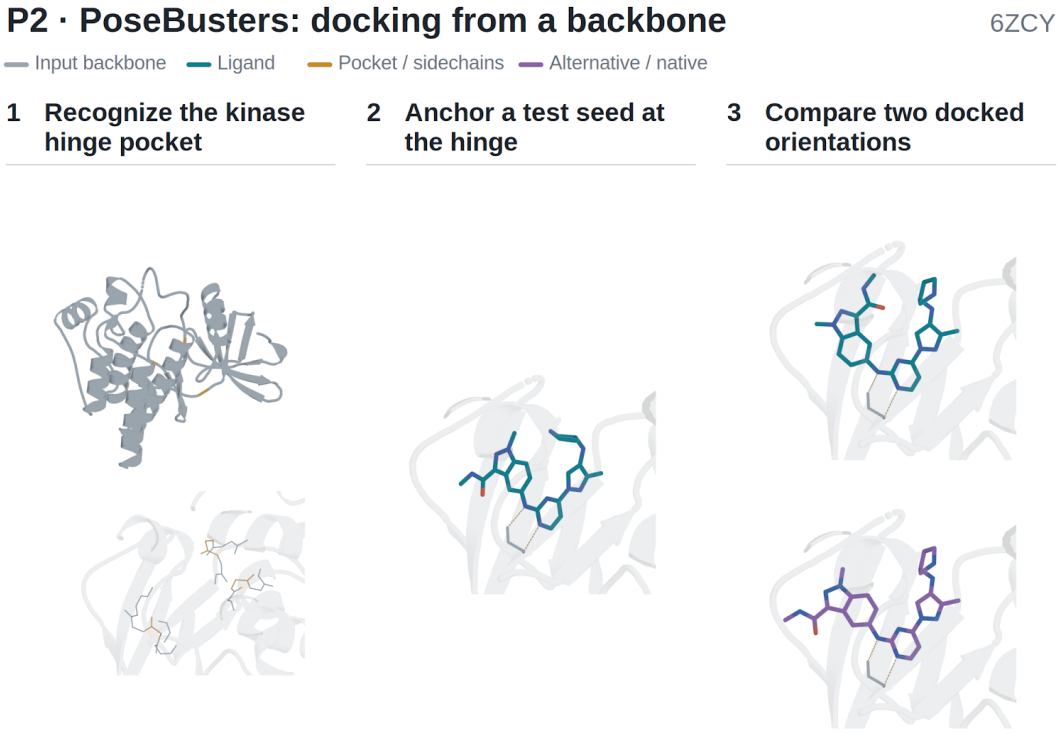
Evaluation part

Submitted Ground truth



Initial molecular views reconstruct backbones; modeled side-chain distances are trace measurements.

E.3.4 Blind Docking

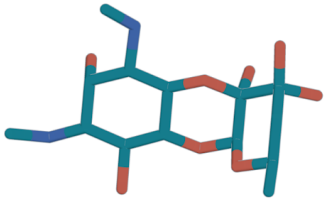


P3 · Build a ligand pose at a site with two metals

7UY4

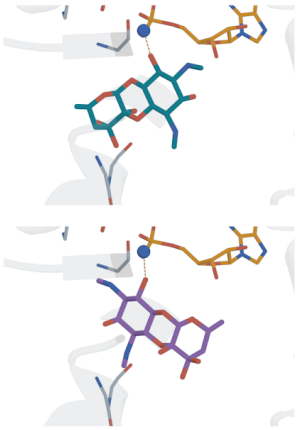
Input backbone — Ligand — Sidechains / ANP — Alternative / native — Mg

1 Build a neutral ligand conformer



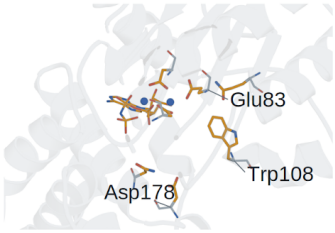
RDKit sampled neutral ligand shapes; MMFF lowered their internal energy while keeping the same bonds.

2 Search yields two ligand orientations



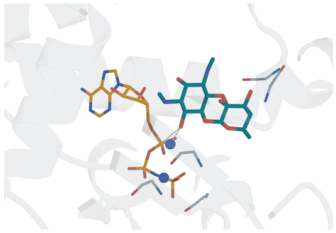
A separate search moved the doubly protonated ligand and sidechains; two cluster means show alternative poses near Mg.

3 Pack seven sidechains together



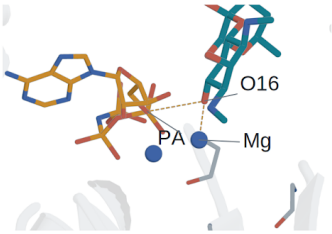
The model ranked sidechain candidates by agreement with sampled structures, then packed seven together to reduce clashes.

4 Inspect a late ligand snapshot



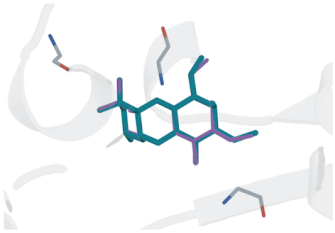
The model checked a late pose for metal contacts, strain and clashes; final coordinates were retrieved again before polishing.

5 Polish and write the final ligand



The model tightened metal restraints, set an O16–PA distance target, and minimized locally before saving the submitted ligand.

Evaluation



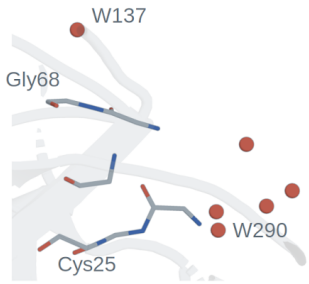
After submission, the grader compared the ligand with the native pose in the backbone frame: RMSD 0.210 Å.

S2 · Scarpino: docking while rebuilding sidechains

5MAJ

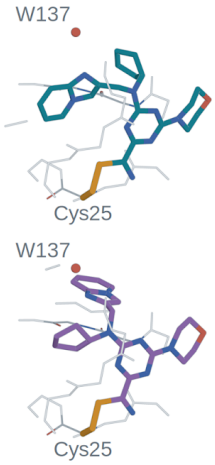
Backbone — Ligand — Sidechains / attachment — Alternative / native — Water

1 Start from backbone and fixed waters



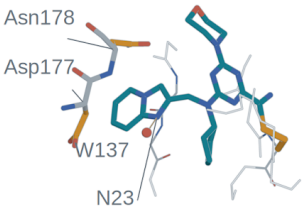
Backbone atoms and fixed waters define the pocket; missing sidechains must be rebuilt around the Cys25 attachment site.

2 Compare two covalent docking alternatives



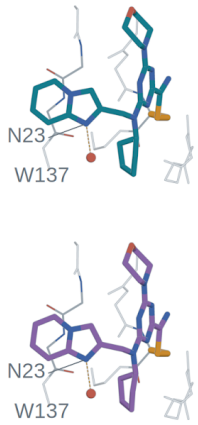
Two separate docking alternatives keep the covalent proxy near Cys25 while placing the ligand substituents differently.

3 Reoptimize and repack the uncertain neighbor



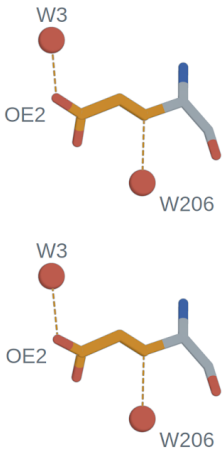
Hold the ligand fixed while rebuilding Asp177/Asn178 in an inferred neighbor; the ligand's original parent was not retained.

4 Refine the ligand with sampled pose information



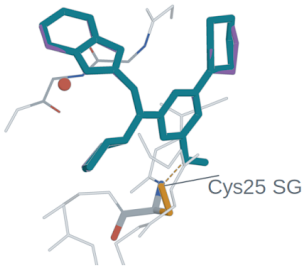
Refine the selected pose locally; a separate later branch uses sampled pose means as restraints for another selection.

5 Repair the crowded Glu190 water contact



Repair Glu190 from the crowded state: its OE2–W3 distance rises from 2.18 to 2.44 Å while the backbone stays fixed.

Evaluation



Submitted ligand versus native in the shared frame: RMSD 0.141 Å after symmetry matching; final Cys25 proxy shown.

S3 · Scarpino: refine a covalent pocket pose

5F2E

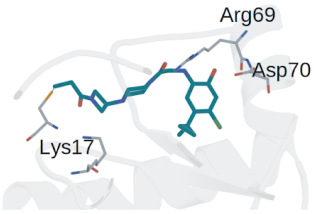
Protein / seed · Ligand · Anchor / contacts · Alternative / native

1 Start from the supplied covalent anchor



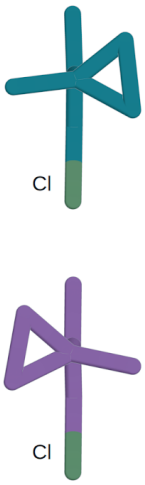
The supplied Cys13 sulfur defines the covalent attachment; nearby pocket geometry guides the model's placement.

2 Place a charged candidate across the pocket



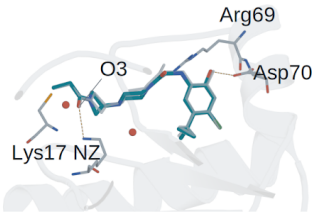
Landmark fitting and local optimization place charged candidate 1286 from Cys13 toward the distal aromatic group.

3 Compare a competing cyclopropyl rotamer



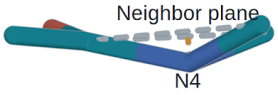
Fragments from separate charged candidates show −86° and +81° torsions, pointing the methyl group in opposite directions.

4 Refine candidate 1286 with molecular mechanics



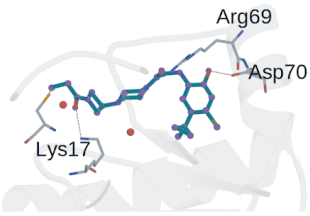
MM continues candidate 1286, adjusting ligand and nearby hydrogens against fixed heavy atoms; O3–Lys17 NZ is 2.93 Å.

5 Inspect the amide distortion after QM/MM



The charged QM/MM snapshot has a less planar amide (N4 height 0.373 vs 0.167 Å in MM); an unprinted MM step lies between.

Evaluation



A separate neutral branch produced the submitted adduct pose on the input graph; native ligand RMSD 0.090 Å.