OMNIPEPBUILDER: UNIFIED PEPTIDE BINDER DESIGN VIA SYSTEMATIC EXPLORATION OF BINDING LANDSCAPES

Anonymous authors
Paper under double-blind review

ABSTRACT

Peptides offer a versatile modality for modulating protein–protein interactions, but scarce complex structures constrain design across diverse cyclization chemistries. Training-free methods repurpose binding-related knowledge encoded in pretrained structure predictors, yet peptide structure prediction and scoring remain insufficiently benchmarked, and efficient exploration of diverse poses remains challenging on difficult targets. We begin with rigorous structural and binding benchmarks to improve prediction and reassess scoring, finding that confidence and Rosetta scores are more informative in controlled mutation series than in heterogeneous screening. We therefore introduce OmniPepBuilder, a unified, training-free two-stage framework for linear and cyclic peptide binder design. To address limited pose diversity and sampling efficiency, global proposal combines state replay, dynamic structural cell partitioning and revisit penalties for adaptive, systematic exploration of binding basins. Local refinement then uses physics-based scoring to optimize interfaces within selected basins. Computational benchmarks show pronounced gains in pose discovery on difficult targets, while local refinement further improves predicted interactions in already confident complexes. Encouraged by this exploration capability, we extend the framework to multi-target binders and macrocyclic molecular glues, where a single peptide must satisfy multiple interface constraints. The putative designs accommodate distinct target pockets in separate complexes or bridge protein partners in diverse ternary poses, demonstrating efficient exploration under the joint geometric constraints of these demanding applications.

1 INTRODUCTION

Binder design sampling repurposes the structural knowledge encoded in protein structure predictors. Generative models learn to sample favorable structural regions, a view supported by the statistical-potential interpretation of protein diffusion models (Roney et al., 2025). RFdiffusion transfers this knowledge from a prediction checkpoint (Watson et al., 2023), BoltzGen learns it through joint prediction–design training (Stark et al., 2025), and the flow-matching model Proteina-Complexa draws on synthetic structural data (Didi et al., 2026). Training-free approaches instead use prediction confidence directly as a pseudo-energy for optimization (Roney & Ovchinnikov, 2022). White-box methods, including BindCraft (Pacesa et al., 2025) and BoltzDesign1 (Cho et al., 2025a), differentiate through the predictor, whereas black-box methods, including BAGEL (Lála et al., 2025) and EvoBind2 (Li et al., 2025a), use its outputs to guide sequence search. Recent diffusion hallucination methods, including Protein Hunter (Cho et al., 2025b) and HalluDesign (Fang et al., 2025), alternate structure prediction from initially uninformative sequences with structure-conditioned sequence design using inverse-folding models such as ProteinMPNN (Dauparas et al., 2022). These Gibbs-like updates explore favorable regions of sequence–structure space through stochastic transitions (Fig. 1a). Related prediction-and-design iterations also enable zero-shot design of small-molecule-binding proteins (Fry et al., 2026).

Scoring complements sampling by translating noisy model outputs into candidate rankings and selection criteria through a minimal set of robust metrics and thresholds. For minibinders, large-scale, consistent experimental binding measurements across targets of interest (Cao et al., 2022) have en-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

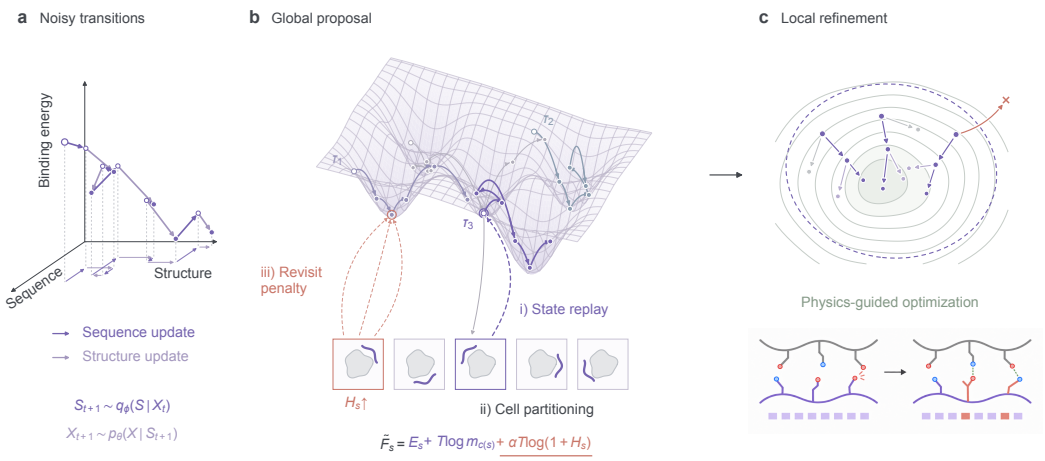


Figure 1: OmniPepBuilder couples noisy alternating sequence–structure sampling (a) with global proposal through state replay, cell partitioning, and revisit penalties (b), followed by physics-guided local refinement within a selected binding basin (c). Landscapes and trajectories are schematic.

abled systematic benchmarking and candidate filtering (Zambaldi et al., 2024; Team et al., 2025; Overath et al., 2025). These empirically grounded criteria support increasingly automated design with reduced human intervention (Claude Science & Shanehsazzadeh, 2026), while also providing objectives for computational search and in silico evaluation.

These foundations are less established for peptide binder design, particularly across noncanonical chemistries and cyclization topologies. Even canonical peptide complexes challenge structure prediction (Zhai et al., 2025), and systematic benchmarking across diverse topologies remains limited (Section 2.1). Scarce complex structures also constrain learning of peptide binding landscapes. Existing approaches address this scarcity through synthetic data and structural augmentation, as in DiffPepBuilder (Wang et al., 2024), CPSea (Yang et al., 2025), and PepGLAD (Kong et al., 2024), or knowledge transfer across molecular types, from antibody interfaces in PepMimic (Kong et al., 2026b) to broader representations in UniMoMo (Kong et al., 2025) and AnewOmni (Kong et al., 2026a). Despite binding datasets for peptide-binding domains (PBDs) (Cunningham et al., 2020) and recent peptide binder screens (Wang et al., 2025a), scoring still lacks broadly validated metrics and filters across therapeutic targets. In particular, interface confidence does not reliably resolve differences in peptide binding affinity (Trolliet et al., 2026).

These gaps leave substantial room to improve both sampling and scoring, particularly with training-free approaches built on generalist structure predictors such as AlphaFold 3 (Abramson et al., 2024), Protenix-v2 (Zhang et al., 2026b), Boltz-2 (Passaro et al., 2025), and OpenDDE (Aureka AI OpenDDE project, 2026), which encode structural knowledge across molecular types. We begin by rigorously benchmarking representations across peptide topologies to improve the predictions underlying sampling (Section 2.1). We then reassess binding scores using heterogeneous PPIKB measurements and controlled mutation series (Section 2.2). Confidence and Rosetta scores are more informative in the latter, while pose recovery itself carries binding information. These findings motivate OmniPepBuilder, a two-stage framework that separates global discovery of binding basins from local optimization within them (Section 3; Fig. 1).

To address inefficient exploration and limited pose diversity on difficult targets, we develop a Go-Explore-inspired global proposal strategy (Ecoffet et al., 2021). It combines state replay, dynamic structural cell partitioning and revisit penalties to preserve promising intermediates and adaptively allocate computation across binding basins (Section 3.1). Subsequent local refinement confines sequence search to a selected basin and uses physics-based scoring as complementary guidance beyond predictor confidence (Section 3.2). Computational benchmarks show pronounced gains over independent trajectories in discovering diverse acceptable poses on difficult targets, while local refinement improves predicted interactions in already confident complexes. Encouraged by this exploration capability, we extend the framework to multi-target binders and macrocyclic molecular glues, where one sequence must satisfy multiple interface constraints (Section 4). The resulting pu-

tative candidates accommodate distinct target pockets in separate complexes or bridge two proteins through diverse ternary poses.

We summarize our contributions as follows:

- We benchmark structural prediction and binding scoring to reassess common practices in peptide design. More explicit cyclization encoding can impair structural fidelity, while scores weak in heterogeneous screening can still inform local sequence optimization. These findings guide structural representations and motivate the two-stage framework (Section 2).
- We introduce OmniPepBuilder with an adaptive strategy for systematic exploration of peptide binding landscapes. State replay, structural cells and revisit penalties yield pronounced gains on difficult targets, complemented by physics-guided optimization within selected basins (Section 3).
- We extend the framework to multi-target binders and macrocyclic molecular glues, generating putative candidates that accommodate distinct target pockets or bridge protein partners in diverse ternary poses (Section 4).

2 Preliminary Explorations

2.1 Structural representations across peptide topologies

Selecting cyclization representations for structure-based peptide design requires benchmarks with verified peptide identities, covalent connectivity and biological binding contexts. Such benchmarks remain scarce beyond head-to-tail amides, leaving suitable representations unsettled across topologies. Released AlphaFold 3 (AF3) preprocessing does not preserve declared polymer–polymer cyclization bonds (Abramson et al., 2024). To adapt predictors to cyclic peptides, HighFold introduced cyclic relative positional encoding (RPE) for head-to-tail and disulfide connections (Zhang et al., 2024), but its complex benchmarks concentrate on related targets and scaffolds, with recurrent binding modes and mixed topologies (Appendix A.1). Protenix, meanwhile, supports explicit connectivity and token-pair bond features (Zhang et al., 2026c;b). These differences call for direct comparison across chemistries and predictors.

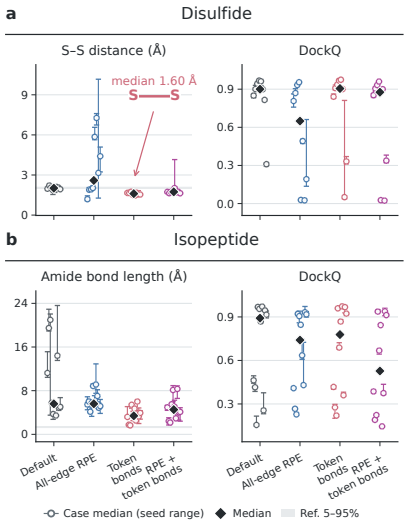


Figure 2: Structural input benchmarking across eight disulfide (a) and ten isopeptide (b) complexes, comparing bond distances and DockQ for four input representations.

To this end, we curated 26 peptide–protein complexes with peptide lengths of 5–30 residues by combining target-sequence clustering at 40% identity with manual structural inspection: eight head-to-tail, eight disulfide, and ten isopeptide complexes. We retained standalone bound peptides with verified cyclization chemistry and adapted AF3 and Protenix v1/v2 to preserve their connectivity. We then compared native residue offsets, topology-aware RPE, token-bond features, and their combination, evaluating both binding-pose recovery and local covalent geometry (Appendix A).

For head-to-tail peptides, cyclic RPE supports accurate backbone closure and greater robustness to the arbitrary sequence origin, motivating its use as the default representation (Fig. 10 and Fig. 11). Unexpectedly, extending RPE or token-bond features to disulfide links can impair predictions: all-edge RPE disrupts bond geometry and degrades pose recovery in several cases, while token-bond features in particular systematically compress S–S distances (Fig. 2a). We therefore preserve disulfide connectivity without these additional features. Isopeptide responses are more heterogeneous. Unmodified positional and token features already recover favorable binding poses, whereas token-bond features bring the linked endpoints closer (Fig. 2b). We use the latter as a closure cue for isopeptide design.

Across the three predictors, these topology-specific choices support broadly comparable head-to-tail pose recovery, while AF3 performs strongly on disulfides and isopeptide complexes (Fig. 12). We consequently use AF3 for global proposal with cyclic RPE for head-to-tail peptides, native features for disulfides, and token-bond hints for isopeptides. Appendix A details the benchmark, implementation adaptations, and full results.

2.2 Binding prediction within and across pose basins

Predicting peptide binding affinity remains challenging (Trolliet et al., 2026). Our benchmark of 368 peptides across 17 PPIKB target contexts (Zhu et al., 2026) likewise finds weak agreement between measured affinity and AF3 interface confidence or Rosetta interface energy (Fig. 3a). However, across different predicted poses, these scores reflect both recovery of a plausible binding mode and the favorability of interactions within it. We therefore hypothesize that establishing a supported pose basin makes sequence comparisons more informative. This perspective parallels Rosetta mutation-effect protocols that evaluate local structural accommodation around a known complex (Barlow et al., 2018).

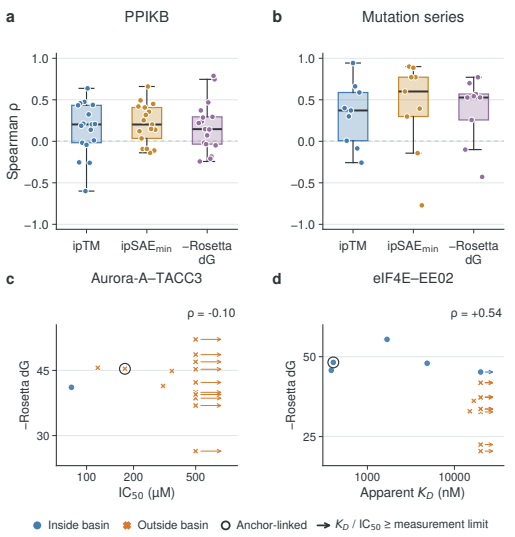


Figure 3: Agreement of confidence and Rosetta scores with literature binding measurements across 17 PPIKB contexts (a) and nine mutation series (b), with Aurora-A–TACC3 and eIF4E–EE02 examples (c, d). Correlations exclude numerical bounds.

To test whether controlled sequence comparisons reveal stronger binding signals, we curated nine mutation series with comparable assays and experimentally grounded pose references. In contrast to their weak screening performance, the same metrics yield stronger correlations across 104 exact measurements (Fig. 3b): median Spearman ρ rises from 0.20 to 0.37 for ipTM, 0.20 to 0.60 for ipSAE$_{\min}$, and 0.15 to 0.53 for negative Rosetta dG. Interface-focused ipSAE (Dunbrack, 2025) is particularly informative in several series, consistent with evidence from designed protein binders (Overath et al., 2025). Appendix B gives curation, protocols, and full results.

Pose recovery itself also carries binding information. For Aurora-A–TACC3, the sole prediction within the reference basin corresponds to the strongest measured binder (Fig. 3c). For eIF4E–EE02, all 11 predictions outside the basin correspond to weak binders, including nine with apparent $K_D > 20\,\mu$M (Fig. 3d). These observations motivate the two-stage framework developed in Section 3: global proposal establishes structurally supported binding modes, while local refinement preserves basin membership and evaluates changes in interfacial interactions (Fig. 1c).

3 OmniPepBuilder

3.1 Systematic Exploration of Binding Basins via Global Proposal

Diffusion-based hallucination alternates structure prediction and sequence design over a heterogeneous binding landscape. Targets differ in difficulty, and initialization and stochastic transitions produce variable outcomes even for the same target. Fixed-length trajectories can discard useful intermediates before yielding acceptable designs. Extending every trajectory does not restore earlier opportunities; our HalluDesign baseline already uses 30 cycles (Fang et al., 2025). With costly, stochastic predictors such as AF3 (Abramson et al., 2024), the central allocation problem is which observed states deserve further exploration.

This loss of useful intermediates resembles detachment in sparse-reward reinforcement learning, which Go-Explore addresses by remembering, returning to and exploring from promising states

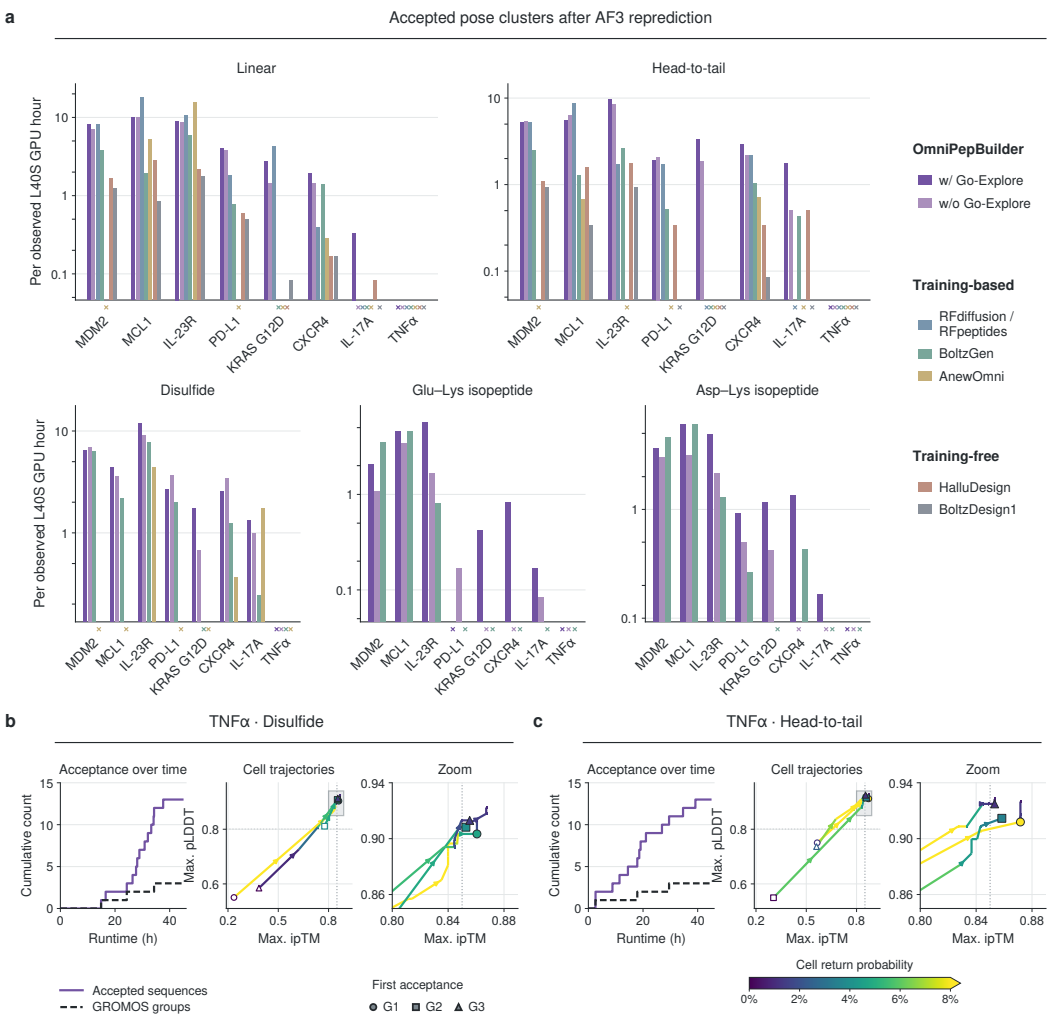


Figure 4: Global proposal benchmarking by common AF3 reprediction (a), showing sampling-adjusted pose-group throughput under a hypothetical linear-growth assumption, and native TNF-alpha search trajectories for disulfide (b) and head-to-tail (c) peptides. Benchmark costs use GPU-hours; trajectory axes use elapsed execution hours on four GPUs.

(Ecoffet et al., 2021). Intuitively, we restore an archived sequence–structure state and sample a new trajectory from it. Such replay allocates variable total exploration depth through short rollouts, revisiting promising states without rediscovering them from random sequences. Compared with rollout-based Monte Carlo tree search, direct replay avoids repeated terminal lookahead to estimate intermediate states' future value through costly structure predictions.

To limit redundancy as the restart pool grows, we retain local Pareto alternatives in interface confidence and peptide confidence, then partition them into structural cells using GROMOS clustering (Daura et al., 1999). Pose similarity is measured in the target reference frame, accounting for peptide length, cyclic registration and target symmetry. Cells are dynamically recomputed as the retained pool changes, while every retained state remains available for replay. To focus computation on states approaching acceptance, we assign a Boltzmann weight based on a confidence-derived pseudo-energy. However, we observed that prioritizing quality alone repeatedly concentrates search in readily accessible regions. Inspired by history-dependent bias in enhanced sampling (Laio & Parrinello, 2002), we penalize repeated returns to similar poses, progressively shifting relative priority toward less explored alternatives (Fig. 1b).

Let $\mathcal{V}$ be the pose history and $\mathcal{S}$ the retained replay pool. For $s \in \mathcal{S}$, let I_s and P_s denote binder ipTM and peptide pLDDT on a 0–1 scale. We define

$$E_s = \max\{0,\, \theta_I - I_s,\, \theta_P - P_s\}, \qquad H_s = \sum_{u \in \mathcal{V}} K(s,u) R_u, \tag{1}$$

where $(\theta_I, \theta_P) = (0.85, 0.80)$, $K(s,u)$ indicates direct pose similarity, and R_u counts return allocations from historical source u. Conditional on replay, the probability of selecting $s \in \mathcal{S}$ is

$$p(s \mid \text{return}) = \frac{w_s}{\sum_{u \in \mathcal{S}} w_u}, \qquad w_s = \frac{\exp(-E_s/T)}{m_{c(s)}(1+H_s)^{\alpha}}, \tag{2}$$

with $m_{c(s)}$ the number of retained states in its current cell, $T = 0.10$ and $\alpha = 0.5$. Cell-size normalization prevents a densely represented region from gaining priority solely through multiplicity; shared return counts preserve exploration pressure as the archive changes. Equivalently, the allocation energy is $\widetilde{F}_s = E_s + T \log m_{c(s)} + \alpha T \log(1 + H_s)$. A fresh origin is drawn with probability 0.30, or whenever the replay pool is empty. Algorithm 1 summarizes the search; Appendix C.1 and Appendix C.2 give implementation and infrastructure details.

Algorithm 1 Global proposal with pose-aware replay

Require: Target and topology; compute budget B; rollout cap $L = 8$; return probability $\rho = 0.70$; $T = 0.10$; $\alpha = 0.5$.

```
Initialize pose history V, match graph K, return counts R, accepted set A.
while budget or dispatched work remains do
    Commit completed evaluations; update A and register eligible poses in V, K.
    S ← LocalPareto(V, K); C ← GROMOS(S, K).
    while origin budget and lineage capacity remain do
        if S = ∅ or u ~ U(0,1) ≥ ρ then
            Initialize a fresh random sequence with a resolved topology.
        else
            Draw s ~ p(s | return) using Eq. 2.
            Charge R_s ← R_s + 1; restore s and sample a ProteinMPNN child.
        end if; start the new lineage at local cycle 1.
    end while
    Advance eligible lineages; stop at acceptance, rejection, duplication, or L evaluations.
end while; return A.
```

We benchmark global proposal across eight targets and five peptide topologies against generative and hallucination methods (Fig. 4a; Appendix C.3). Under common AF3 reprediction, OmniPepBuilder attains the highest positive pose-group throughput in 21 of 40 target–topology cases. Gains are particularly evident in challenging settings: head-to-tail designs yield 40 accepted groups for KRAS G12D, 35 for CXCR4 and 21 for IL-17A, compared with 22, 26 and 6 without replay.

Under Protenix-v2 reprediction, OmniPepBuilder retains passing designs in 21 of 24 linear, head-to-tail and disulfide cases, although method rankings change (Fig. 14). We argue that this is a test of successive predictor-specific selection: our candidates already pass native AF3 acceptance, and concordance with Protenix's distinct confidence landscape need not imply stronger binding. Protenix also yields very few passing candidates for the hardest topologies: only two of 9,616 isopeptide candidates pass the full evaluation, despite target templates, ten recycling cycles and five samples. Repredicted confidence and interface-energy distributions are shown in Fig. 15 and Fig. 16.

Our main benchmark yielded no accepted TNFα designs, prompting extended search experiments to test whether further exploration could uncover favorable basins (Fig. 4b,c). Disulfide and head-to-tail searches each identify 13 native accepted sequences spanning three pose groups under stringent confidence, hotspot and topology criteria. We further visualize cell histories to examine how these discoveries arise: progression toward acceptance accompanies declining allocation probabilities in repeatedly explored regions. Repredicted representatives show that the accepted groups capture alternative binding poses (Fig. 18); Appendix C.5 gives the extended-search settings and trajectory measurements.

To further evaluate the combined exploration policy, we compare three ablations (Fig. 5). Removing revisit penalties, spatial cell organization or replay reduces median accepted-group yield by 3.7%, 29.9% and 14.0%, respectively, across 14 cases at 12 generation GPU-hours. Without revisit penalties, some easier cases accumulate accepted groups faster, whereas their benefit is clearer on harder cases (Fig. 17): full-policy yields increase from 26 to 33 for linear KRAS G12D, from 32 to 40 for head-to-tail KRAS G12D, and from 26 to 35 for head-to-tail CXCR4. These results support allocating additional search to underexplored regions while retaining productive states for further expansion.

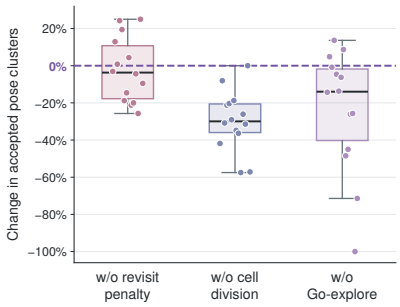


Figure 5: Changes in accepted pose-group counts at 12 generation GPU-hours for three exploration-policy ablations relative to full Go-Explore. The 14 cases per arm exclude two TNF-alpha cases with zero reference yield.

3.2 Physics-Guided Interface Optimization via Local Refinement

The mutation benchmark (Section 2.2) motivates optimizing sequence-dependent interactions within a basin established by global proposal. Having optimized predictor confidence, we preserve the selected binding mode and use Rosetta interface energy as a complementary, physically based signal. Basin confinement is the defining principle: another benchmarked objective, such as $\text{ipSAE}_{\text{min}}$, can replace Rosetta. With validated chemical representations and scoring, this principle could extend to noncanonical amino acids.

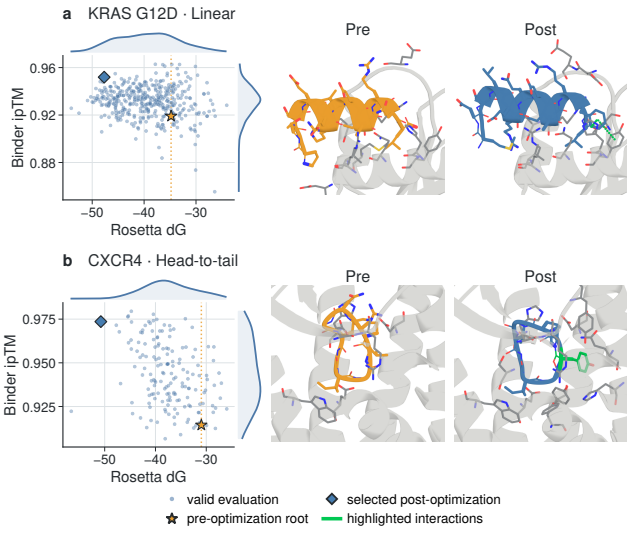


Figure 6: Local refinement of AF3-generated candidates for KRAS G12D (linear, a) and CXCR4 (head-to-tail, b). Scatter plots show Rosetta dG and Protenix ipTM; green structural highlights mark Phe12 aromatic packing and Tyr1 pocket insertion, respectively.

We propose single-residue changes by mixing uniform substitutions with ProteinMPNN conditional probabilities (Dauparas et al., 2022). Each candidate preserves length and topology and differs from the starting sequence at no more than five root-registered positions; equivalent cyclic rotations are deduplicated at output. For a starting sequence s_0 and pose X_0, the optimization is

$$\mathcal{B}_5(s_0) = \{s : d_H(s, s_0) \le 5\}, \qquad \min_{s \in \mathcal{B}_5(s_0),\, G(s;X_0)=1} \big[E(s) - E(s_0)\big], \tag{3}$$

where G requires confident prediction in the reference basin and controlled structural drift during relaxation, and E is Rosetta dG after relaxation. We implemented and evaluated evolutionary search and Metropolis/annealing on six linear starting candidates. After preliminary parameter calibration, evolutionary search achieved lower run-median dG in 15 of 17 paired runs at 192 evaluations per

run, motivating its selection as an empirical default. Appendix D provides the algorithms, settings and comparison.

We further inspect two selected cases to examine how local refinement improves predicted interactions in already confident complexes (Fig. 6). For KRAS G12D, five substitutions reduce dG from -34.80 to -47.69 REU; the introduced Phe12 forms π–π packing interactions with His95 and Tyr96. For head-to-tail CXCR4, three substitutions reduce dG from -30.97 to -50.76 REU, with Tyr1 occupying the hydrophobic pocket more extensively than the original Leu1. Protenix ipTM remains high, increasing from 0.919 to 0.952 and from 0.914 to 0.974, respectively. These examples illustrate how a physical objective can guide local interaction improvements beyond the confidence criteria used to discover the initial basin.

4 Designing Multi-target Binders and Macrocyclic Glues

Encouraged by efficient exploration of single-target binding poses, we extend global proposal to multi-target peptides, which can coordinate complementary pharmacology within one molecule. Dual KRAS G12D/HDAC2 inhibition couples oncogenic signaling with epigenetic regulation (Zhang et al., 2026a), while MCL-1/BCL-xL co-inhibition addresses compensatory cancer-survival pathways (Wang et al., 2025b). Retatrutide demonstrates the therapeutic potential of metabolic polyagonism: it combines GLP1R, GCGR and GIPR activity and achieved substantial weight loss in a phase 2 trial (Coskun et al., 2022; Jastreboff et al., 2023). Explicitly designing such activity remains challenging; GLP1R activity, for example, emerged serendipitously from EvoBind peptides designed solely against GCGR (Li et al., 2025b).

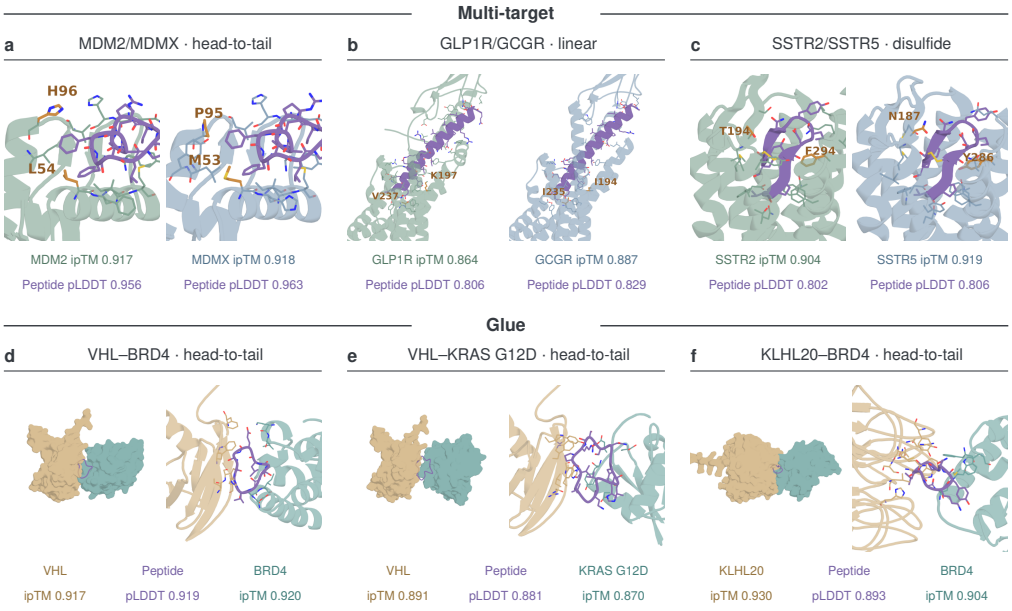


Figure 7: Multi-target peptide candidates in separate MDM2/MDMX, GLP1R/GCGR and SSTR2/SSTR5 complexes (a–c), and head-to-tail macrocyclic glue candidates bridging VHL–BRD4, VHL–KRAS G12D and KLHL20–BRD4 (d–f). Amber highlights mark differing target-pocket residues; all six candidates pass the computational acceptance criteria.

A complementary strategy uses macrocyclic glues to induce protein proximity, enabling E3-mediated degradation (Jing et al., 2025) or ternary-complex inhibition, exemplified by cyclophilin A recruitment to active RAS by daraxonrasib (Cregg et al., 2025). Recent bifunctional protein designs (Mylemans et al., 2026) and concurrent work using EvoBind-multimer to design macrocyclic glues recruiting VHL to KRAS or BRD4 (Brunner et al., 2026) illustrate this opportunity. However, satisfying multiple interfaces imposes a demanding geometric constraint: a shared sequence must accommodate distinct target surfaces, and a glue must additionally engage both partners in one com-

patible assembly. These intersecting constraints motivate systematic exploration of the restricted set of eligible binding basins.

As a proof of concept, we apply global proposal to therapeutically relevant target combinations, representing multi-target binders in separate target complexes and glues in joint ternary complexes. Pose matching, Pareto retention and acceptance consider every required interface, with replay allocating exploration across joint pose basins (Appendix E).

The MDM2/MDMX macrocycle accommodates both binding pockets (Fig. 7a): Phe4 and Phe5 preserve close hydrophobic contacts in the differing His96/Pro95 and Leu54/Met53 environments. This shared anchoring pattern shows how one compact scaffold can satisfy two nonidentical interfaces. Linear GLP1R/GCGR and disulfide-constrained SSTR2/SSTR5 candidates extend this compatibility across peptide topologies (Fig. 7b,c).

We next test simultaneous engagement of both partners in a ternary assembly. The glue macrocycles bridge VHL–BRD4, VHL–KRAS G12D and KLHL20–BRD4 through compact contacts with both protein surfaces (Fig. 7d–f). Search recovers alternative peptide placements and partner orientations without a preassembled complex: VHL–BRD4 and KLHL20–BRD4 yield 19 and six accepted pose groups, respectively. All campaigns ran for less than 24 hours on four L40S GPUs, demonstrating efficient discovery of diverse solutions under joint interface constraints. Appendix E reports additional structures and campaign outcomes.

5 Related Work

Target-specific peptide binder design. Peptides bridge small-molecule and protein chemical space (Quartararo et al., 2020). Adaptations of protein-design methods include RFpeptides (Rettie et al., 2025b) and AfCycDesign (Rettie et al., 2025a), which introduce cyclic positional encodings, while DiffPepBuilder (Wang et al., 2024), PPFlow (Lin et al., 2024) and PepFlow (Li et al., 2024a) learn sequence–structure distributions through backbone or torsional representations. Molecular approaches instead model atomic interactions with PocketXMol (Peng et al., 2026) or residue/fragment latent representations with PepGLAD (Kong et al., 2024), UniMoMo (Kong et al., 2025) and AnewOmni (Kong et al., 2026a). More recently, all-atom prediction architectures have informed generative binder design in PXDesign (Team et al., 2025) and BoltzGen (Stark et al., 2025), alongside all-atom generators such as Latent-X (Latent Labs Team et al., 2025).

Inference-time scaling for binder design. Diffusion search allocates additional compute to noise-space optimization and reward-guided resampling (Ma et al., 2025; Singhal et al., 2025). Proteina-Complexa combines an atomistic generative prior with beam search, Feynman–Kac steering and MCTS, optimizing confidence and interface hydrogen bonding (Didi et al., 2026). Oracle-budget benchmarking highlights the dependence of guidance performance on evaluation budget (Kalisz et al., 2026). Concurrent work includes MCTD-ME, which uses multi-expert tree search over diffusion states (Liu et al., 2026a), and MCTH, which applies uncertainty-aware MCTS to frozen folding and inverse-folding models (Liu et al., 2026b). MCTH's branch-wise statistics do not pool revisits to equivalent poses across trajectories, and its benchmarks do not address diverse cyclic-peptide topologies. OmniPepBuilder instead shares replay statistics across structural pose neighborhoods and separates basin discovery from optimization within a selected basin.

6 Conclusion

We present OmniPepBuilder, a unified, training-free framework for linear and cyclic peptide binder design. Guided by structural and mutation benchmarks, it separates systematic discovery of binding basins from physics-guided optimization within them. Pose-aware replay, structural cell partitioning and revisit penalties improve global exploration, while basin-constrained local refinement captures complementary interaction signals beyond predictor confidence. Computational benchmarks and structural analyses demonstrate competitive design performance across targets and topologies, diverse binding poses, and improved interface interactions. Extensions to multi-target binders and macrocyclic glue candidates illustrate a common search framework for both independent target recognition and simultaneous ternary assembly.

AI USE STATEMENT

We used OpenAI Codex to aid and polish writing and to support research ideation and execution. Research assistance included literature search and synthesis, refinement of research arguments, code inspection, and interpretation of computational results. In particular, Codex with GPT-5.6 Sol and GPT-6 Astra supported the iterative retrieval and curation of primary-literature mutation series, including extraction of peptide sequences and binding measurements, checks of assay and chemical context, and organization of source-linked records (Appendix B.1). Codex assisted in drafting selected paragraphs. The authors are responsible for finalizing the manuscript and for its final content.

REPRODUCIBILITY STATEMENT

Methodological details. OmniPepBuilder performs training-free search with pretrained structure-prediction and sequence-design models. Section 3 presents the two-stage framework; Appendix C specifies pose matching, Pareto retention, replay allocation, revisit penalties and asynchronous execution. Appendix D provides the local-refinement algorithms, mutation constraints, structural gates and cyclic-sequence deduplication. The joint pose definitions and search adaptations for multi-target binders and molecular glues are described in Appendix E.1.

Data curation and processing. Appendix A.1 documents structural-benchmark selection, peptide topologies, PDB identifiers and complex preparation. Appendix B.1 describes PPIKB filtering, target-context assignment, measurement aggregation and primary-literature mutation-series curation, including a source table, assay endpoints and treatment of bounded measurements. The design-benchmark targets, prepared residue ranges and hotspots are listed in Appendix C.3.

Evaluation and computational settings. Appendix A and Appendix B report predictor settings, sample selection, structural and binding metrics, and statistical procedures. Appendix C.3 specifies baseline implementations and checkpoints, sampling parameters, common reprediction, acceptance criteria, compute accounting and ablation settings. Local-refinement hyperparameter comparisons and evaluation budgets are given in Appendix D.2; application-specific filters, hardware and campaign budgets appear in Appendix E.1.

Code release. Upon acceptance, we will publicly release the OmniPepBuilder source code, including scripts and configuration files for benchmark preparation, evaluation and analysis.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024. doi: 10.1038/s41586-024-07487-w.

Jeanine F. Amacher, Patrick R. Cushing, Lionel Brooks, Prisca Boisguerin, and Dean R. Madden. Stereochemical preferences modulate affinity and selectivity among five pdz domains that bind cftr: Comparative structural and sequence analyses. *Structure*, 22(1):82–93, January 2014. ISSN 0969-2126. doi: 10.1016/j.str.2013.09.019. URL http://dx.doi.org/10.1016/j.str.2013.09.019.

Aureka AI OpenDDE project. Folding, Reasoning, and Scaling with Open-source Drug Discovery Engine. *arXiv preprint arXiv:2607.03787*, 2026. URL https://arxiv.org/abs/2607.03787.

Kyle A. Barlow, Shane Ó Conchúir, Samuel Thompson, Pooja Suresh, James E. Lucas, Markus Heinonen, and Tanja Kortemme. Flex ddg: Rosetta ensemble-based estimation of changes in

protein–protein binding affinity upon mutation. *The Journal of Physical Chemistry B*, 122(21): 5389–5399, February 2018. ISSN 1520-5207. doi: 10.1021/acs.jpcb.7b11367. URL http://dx.doi.org/10.1021/acs.jpcb.7b11367.

Andrä Brunner, Krzysztof Wierbilowicz, Diandra Daumiller, Daniel Bexell, Kasper Karlsson, Olle Sangfelt, and Patrick Bryant. De novo Design of Macrocyclic Molecular Glues. *bioRxiv*, August 2026. doi: 10.64898/2026.08.21.746227. URL http://dx.doi.org/10.64898/2026.08.21.746227.

Longxing Cao, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Joon Sung Park, Kevin M. Jude, Iva Marković, Rameshwar U. Kadam, Koen H. G. Verschueren, Kenneth Verstraete, Scott Thomas Russell Walsh, Nathaniel Bennett, Ashish Phal, Aerin Yang, Lisa Kozodoy, Michelle DeWitt, Lora Picton, Lauren Miller, Eva-Maria Strauch, Nicholas D. DeBouver, Allison Pires, Asim K. Bera, Samer Halabiya, Bradley Hammerson, Wei Yang, Steffen Bernard, Lance Stewart, Ian A. Wilson, Hannele Ruohola-Baker, Joseph Schlessinger, Sangwon Lee, Savvas N. Savvides, K. Christopher Garcia, and David Baker. Design of protein-binding proteins from the target structure alone. *Nature*, 605(7910):551–560, 2022. doi: 10.1038/s41586-022-04654-9. URL https://doi.org/10.1038/s41586-022-04654-9.

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

Yehlin Cho, Griffin Rangel, Gaurav Bhardwaj, and Sergey Ovchinnikov. Protein Hunter: exploiting structure hallucination within diffusion for protein design. *bioRxiv*, 2025b. doi: 10.1101/2025.10.10.681530. URL https://doi.org/10.1101/2025.10.10.681530. Preprint.

Claude Science and Amir Shanehsazzadeh. Autonomous de novo protein binder design with Claude. Technical report, Anthropic, August 2026. URL https://www-cdn.anthropic.com/30bf50e22a01388bb29bf077ee3f244531594b7a.pdf.

Tamer Coskun, Shweta Urva, William C. Roell, Hongchang Qu, Corina Loghin, Julie S. Moyers, Libbey S. O'Farrell, Daniel A. Briere, Kyle W. Sloop, Melissa K. Thomas, Valentina Pirro, David B. Wainscott, Francis S. Willard, Matthew Abernathy, LaRonda Morford, Yu Du, Charles Benson, Ruth E. Gimeno, Axel Haupt, and Zvonko Milicevic. LY3437943, a novel triple glucagon, GIP, and GLP-1 receptor agonist for glycemic control and weight loss: From discovery to clinical proof of concept. *Cell Metabolism*, 34(9):1234–1247.e9, September 2022. ISSN 1550-4131. doi: 10.1016/j.cmet.2022.07.013. URL http://dx.doi.org/10.1016/j.cmet.2022.07.013.

James Cregg, Anne V. Edwards, Stephanie Chang, Bianca J. Lee, John E. Knox, Aidan C. A. Tomlinson, Abby Marquez, Yang Liu, Rebecca Freilich, Naing Aay, Yingyun Wang, Lingyan Jiang, Jingjing Jiang, Zhican Wang, Michael Flagella, David Wildes, Jacqueline A.M. Smith, Mallika Singh, Zhengping Wang, Adrian L. Gill, and Elena S. Koltun. Discovery of Daraxonrasib (RMC-6236), a Potent and Orally Bioavailable RAS(ON) Multi-selective, Noncovalent Tri-complex Inhibitor for the Treatment of Patients with Multiple RAS-Addicted Cancers. *Journal of Medicinal Chemistry*, 68(6):6064–6083, March 2025. ISSN 1520-4804. doi: 10.1021/acs.jmedchem.4c02314. URL http://dx.doi.org/10.1021/acs.jmedchem.4c02314.

Joseph M Cunningham, Julia R Rogers, Grigoriy Koytiger, Peter K Sorger, and Mohammed AlQuraishi. Biophysical prediction of protein–peptide interactions and signaling networks using machine learning. *Nature methods*, 17(2):175–183, 2020.

J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker. Robust deep learning–based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022. doi: 10.1126/science.add2187. URL https://doi.org/10.1126/science.add2187.

Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F Van Gunsteren, and Alan E Mark. Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition*, 38(1-2):236–240, 1999.

Kieran Didi, Zuobai Zhang, Guoqing Zhou, Danny Reidenbach, Zhonglin Cao, Sooyoung Cha, Tomas Geffner, Christian Dallago, Jian Tang, Michael M. Bronstein, Martin Steinegger, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Scaling atomistic protein binder design with generative pretraining and test-time compute. In *The Fourteenth International Conference on Learning Representations (ICLR)*, 2026.

Roland L. Dunbrack, Jr. Rēs ipsae loquuntur : What's wrong with alphafold's iptm score and how to fix it. *bioRxiv*, February 2025. doi: 10.1101/2025.02.10.637595. URL http://dx.doi.org/10.1101/2025.02.10.637595.

Adrien Ecoffet, Joost Huizinga, Joel Lehman, Kenneth O. Stanley, and Jeff Clune. First return, then explore. *Nature*, 590(7847):580–586, 2021. doi: 10.1038/s41586-020-03157-9. URL https://doi.org/10.1038/s41586-020-03157-9.

Minchao Fang, Chentong Wang, Jungang Shi, Fengbai Lian, Qihan Jin, Zhe Wang, Yanzhe Zhang, Peipei Chen, Zhanyuan Cui, Yanjun Wang, Ze Zhang, Yitao Ke, Qingzheng Han, and Longxing Cao. HalluDesign: Protein Optimization and de novo Design via Iterative Structure Hallucination and Sequence Design. *bioRxiv*, 2025. doi: 10.1101/2025.11.08.686881. URL https://doi.org/10.1101/2025.11.08.686881. Preprint.

Glenna Wink Foight, Jeremy A. Ryan, Stefano V. Gullá, Anthony Letai, and Amy E. Keating. Designed bh3 peptides with high affinity and specificity for targeting mcl-1 in cells. *ACS Chemical Biology*, 9(9):1962–1968, July 2014. ISSN 1554-8937. doi: 10.1021/cb500340w. URL http://dx.doi.org/10.1021/cb500340w.

Yuri Frosi, Simon Ng, Yen-Chu Lin, Shimin Jiang, Siti Radhiah Ramlan, Dilraj Lama, Chandra S. Verma, Ignacio Asial, and Christopher J. Brown. Development of a novel peptide aptamer that interacts with the eif4e capped-mrna binding site using peptide epitope linker evolution (pele). *RSC Chemical Biology*, 3(7):916–930, 2022. ISSN 2633-0679. doi: 10.1039/d2cb00099g. URL http://dx.doi.org/10.1039/D2CB00099G.

Benjamin Fry, Kaia Slaw, and Nicholas F. Polizzi. Zero-shot design of drug-binding proteins via neural iterative selection-expansion. *Nature*, 656(8126):237–249, 2026. doi: 10.1038/s41586-026-10670-w. URL https://doi.org/10.1038/s41586-026-10670-w.

Diana Gimenez, Martin Walko, Jennifer A. Miles, Richard Bayliss, Megan H. Wright, and Andrew J. Wilson. Constrained tacc3 peptidomimetics for a non-canonical protein–protein interface elucidate allosteric communication in aurora-a kinase. *Chemical Science*, 16(1):354–363, 2025. ISSN 2041-6539. doi: 10.1039/d4sc06100d. URL http://dx.doi.org/10.1039/D4SC06100D.

Aimee J. Horsfall, Beth A. Vandborg, Wioleta Kowalczyk, Theresa Chav, Denis B. Scanlon, Andrew D. Abell, and John B. Bruning. Unlocking the pip-box: A peptide library reveals interactions that drive high-affinity binding to human pcna. *Journal of Biological Chemistry*, 296:100773, January 2021. ISSN 0021-9258. doi: 10.1016/j.jbc.2021.100773. URL http://dx.doi.org/10.1016/j.jbc.2021.100773.

Ania M. Jastreboff, Lee M. Kaplan, Juan P. Frías, Qiwei Wu, Yu Du, Sirel Gurbuz, Tamer Coskun, Axel Haupt, Zvonko Milicevic, and Mark L. Hartman. Triple–Hormone-Receptor Agonist Retatrutide for Obesity — A Phase 2 Trial. *New England Journal of Medicine*, 389(6):514–526, August 2023. ISSN 1533-4406. doi: 10.1056/nejmoa2301972. URL http://dx.doi.org/10.1056/NEJMoa2301972.

Xuefei Jing, Joel P. Mackay, and Toby Passioura. Macrocyclic peptides as a new class of targeted protein degraders. *RSC Chemical Biology*, 6(3):326–337, 2025. ISSN 2633-0679. doi: 10.1039/d4cb00199k. URL http://dx.doi.org/10.1039/D4CB00199K.

Aleksandra Kalisz, Jack Simons, Krisztina Sinkovics, Noam Ghenassia, Shikha Surana, Henry Moss, and Paul Duckworth. How to Spend Your Oracle Budget: Practical Guidance for Protein Structure Prediction Models. arXiv:2608.12192, 2026. URL https://arxiv.org/abs/2608.12192.

Xiangzhe Kong, Yinjun Jia, Wenbing Huang, and Yang Liu. Full-atom peptide design with geometric latent diffusion. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 74808–74839. Curran Associates, Inc., 2024. doi: 10.52202/079017-2379. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/88ad9774ffcb7a272457e9396f793a07-Paper-Conference.pdf.

Xiangzhe Kong, Zishen Zhang, Ziting Zhang, Rui Jiao, Jianzhu Ma, Wenbing Huang, Kai Liu, and Yang Liu. UniMoMo: Unified generative modeling of 3D molecules for de novo binder design. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 31397–31418. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/kong25b.html.

Xiangzhe Kong, Junwei Chen, Ziting Zhang, Gaodeng Li, Qingyuan Zhu, Lei Wei, Mingyu Li, Yan Shi, Weiyang Dai, Zishen Zhang, Wenjuan Tan, Rui Jiao, Xiaolun Wang, Jiqing Zheng, Ziyang Yu, Qilong Wu, Zhiye Guo, Li Zhang, Wentao Li, Qiaojing Huang, Tian Zhu, Xiaowo Wang, Wenbing Huang, Yuli She, Jian Zhang, Yang Liu, Kai Liu, and Jianzhu Ma. Programming Biomolecular Interactions with All-Atom Generative Model. *bioRxiv*, 2026a. doi: 10.64898/2026.03.12.711044. URL https://doi.org/10.64898/2026.03.12.711044. Preprint.

Xiangzhe Kong, Rui Jiao, Haowei Lin, Ruihan Guo, Wenbing Huang, Wei-Ying Ma, Zihua Wang, Yang Liu, and Jianzhu Ma. Peptide design through binding interface mimicry with PepMimic. *Nature Biomedical Engineering*, 10(7):1295–1310, 2026b. doi: 10.1038/s41551-025-01507-4. URL https://doi.org/10.1038/s41551-025-01507-4.

Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the national academy of sciences*, 99(20):12562–12566, 2002.

Latent Labs Team, Alex Bridgland, Jonathan Crabbé, Henry Kenlay, Daniella Pretorius, Sebastian M. Schmon, Agrin Hilmkil, Rebecca Bartke-Croughan, Robin Rombach, Michael Flashman, Tomas Matteson, Simon Mathis, Alexander W. R. Nelson, David Yuan, Annette Obika, and Simon A. A. Kohl. Latent-X: An Atom-level Frontier Model for De Novo Protein Binder Design. arXiv:2507.19375, 2025. URL https://arxiv.org/abs/2507.19375.

Jiahan Li, Chaoran Cheng, Zuofan Wu, Ruihan Guo, Shitong Luo, Zhizhou Ren, Jian Peng, and Jianzhu Ma. Full-Atom Peptide Design based on Multi-modal Flow Matching. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 27615–27640. PMLR, 21–27 Jul 2024a. URL https://proceedings.mlr.press/v235/li24o.html.

Qiuzhen Li, Efstathios Nikolaos Vlachos, and Patrick Bryant. Design of linear and cyclic peptide binders from protein sequence information. *Communications Chemistry*, 8(1):211, 2025a. doi: 10.1038/s42004-025-01601-3. URL https://doi.org/10.1038/s42004-025-01601-3.

Qiuzhen Li, Elisee Wiita, Thomas Helleday, and Patrick Bryant. Blind De Novo Design of Dual Cyclic Peptide Agonists Targeting GCGR and GLP1R. *bioRxiv*, June 2025b. doi: 10.1101/2025.06.06.658268. URL http://dx.doi.org/10.1101/2025.06.06.658268.

Yang Li, Qingtong Zhou, Antao Dai, Fenghui Zhao, Rulue Chang, Tianlei Ying, Beili Wu, Dehua Yang, Ming-Wei Wang, and Zhaotong Cong. Structural analysis of the dual agonism at GLP-1R and GCGR. *Proceedings of the National Academy of Sciences*, 120(33):e2303696120, August 2023. ISSN 1091-6490. doi: 10.1073/pnas.2303696120. URL http://dx.doi.org/10.1073/pnas.2303696120.

Ying-ge Li, Xian-yu Meng, Xiru Yang, Sheng-long Ling, Pan Shi, Chang-lin Tian, and Fan Yang. Structural insights into somatostatin receptor 5 bound with cyclic peptides. *Acta Pharmacologica Sinica*, 45(11):2432–2440, Nov 2024b. ISSN 1745-7254. doi: 10.1038/s41401-024-01314-8. URL https://doi.org/10.1038/s41401-024-01314-8.

Haitao Lin, Odin Zhang, Huifeng Zhao, Dejun Jiang, Lirong Wu, Zicheng Liu, Yufei Huang, and Stan Z. Li. PPFLOW: Target-Aware Peptide Design with Torsional Flow Matching. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 30510–30528. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/lin24z.html.

Xuefeng Liu, Mingxuan Cao, Songhao Jiang, Xiao Luo, Xiaotian Duan, Mengdi Wang, Tobin R. Sosnick, Jinbo Xu, and Rick Stevens. Monte Carlo Tree Diffusion with Multiple Experts for Protein Design. arXiv:2509.15796, 2026a. URL https://arxiv.org/abs/2509.15796.

Xuefeng Liu, Mingxuan Cao, Xiao Luo, Songhao Jiang, Tobin Sosnick, Jinbo Xu, Louis Maher, and Rick Stevens. Leveraging generative hallucination and biophysics-informed modeling for unified biomolecular sequence-structure co-design. arXiv:2608.17381, 2026b. URL https://arxiv.org/abs/2608.17381.

Jakub Lála, Ayham Al-Saffar, and Stefano Angioletti-Uberti. BAGEL: Protein engineering via exploration of an energy landscape. *PLOS Computational Biology*, 21(12):e1013774, 2025. doi: 10.1371/journal.pcbi.1013774.

Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, and Saining Xie. Inference-Time Scaling for Diffusion Models beyond Scaling Denoising Steps. arXiv:2501.09732, 2025. URL https://arxiv.org/abs/2501.09732.

Caleb B. McDonald, Kenneth L. Seldeen, Brian J. Deegan, and Amjad Farooq. Sh3 domains of grb2 adaptor bind to pxψpxr motifs within the sos1 nucleotide exchange factor in a discriminate manner. *Biochemistry*, 48(19):4074–4085, April 2009. ISSN 1520-4995. doi: 10.1021/bi802291y. URL http://dx.doi.org/10.1021/bi802291y.

Claudio Mirabello and Björn Wallner. Dockq v2: improved automatic quality measure for protein multimers, nucleic acids, and small molecules. *Bioinformatics*, 40(10):btae586, 2024. doi: 10.1093/bioinformatics/btae586.

Bram Mylemans, Boguslawa Korona, Amanda M. Acevedo-Jake, Ailsa MacRae, Thomas A. Edwards, Danny T. Huang, Andrew J. Wilson, Laura S. Itzhaki, and Derek N. Woolfson. De Novo-Designed Bifunctional Proteins for Targeted Protein Degradation. *Journal of the American Chemical Society*, 148(32):34311–34319, August 2026. ISSN 1520-5126. doi: 10.1021/jacs.6c07593. URL http://dx.doi.org/10.1021/jacs.6c07593.

Max D. Overath, Andreas S. H. Rygaard, Christian P. Jacobsen, Valentas Brasas, Oliver Morell, Pietro Sormanni, and Timothy P. Jenkins. Predicting Experimental Success in De Novo Binder Design: A Meta-Analysis of 3,766 Experimentally Characterised Binders. *bioRxiv*, 2025. doi: 10.1101/2025.08.14.670059. URL https://doi.org/10.1101/2025.08.14.670059. Preprint.

Martin Pacesa, Lennart Nickel, Christian Schellhaas, Joseph Schmidt, Ekaterina Pyatova, Lucas Kissling, Patrick Barendse, Jagrity Choudhury, Srajan Kapoor, Ana Alcaraz-Serna, Yehlin Cho, Kourosh H. Ghamary, Laura Vinué, Brahm J. Yachnin, Andrew M. Wollacott, Stephen Buckley, Adrie H. Westphal, Simon Lindhoud, Sandrine Georgeon, Casper A. Goverde, Georgios N. Hatzopoulos, Pierre Gönczy, Yannick D. Muller, Gerald Schwank, Daan C. Swarts, Alex J. Vecchio, Bernard L. Schneider, Sergey Ovchinnikov, and Bruno E. Correia. One-shot design of functional protein binders with BindCraft. *Nature*, 646(8084):483–492, 2025. doi: 10.1038/s41586-025-09429-6. URL https://doi.org/10.1038/s41586-025-09429-6.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique

Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025. doi: 10.1101/2025.06.14.659707.

Marzena Pazgier, Min Liu, Guozhang Zou, Weirong Yuan, Changqing Li, Chong Li, Jing Li, Juahdi Monbo, Davide Zella, Sergey G. Tarasov, and Wuyuan Lu. Structural basis for high-affinity peptide inhibition of p53 interactions with MDM2 and MDMX. *Proceedings of the National Academy of Sciences*, 106(12):4665–4670, March 2009. ISSN 1091-6490. doi: 10.1073/pnas.0900947106. URL http://dx.doi.org/10.1073/pnas.0900947106.

Xingang Peng, Ruihan Guo, Fenglin Guo, Ziyi Wang, Jiayu Sun, Jiaqi Guan, Yinjun Jia, Yan Xu, Yanwen Huang, Muhan Zhang, Jian Peng, Xinquan Wang, Chuanhui Han, Zihua Wang, and Jianzhu Ma. Unified modeling of 3D molecular generation via atomic interactions with PocketX-Mol. *Cell*, 189(7):1904–1922.e28, April 2026. ISSN 0092-8674. doi: 10.1016/j.cell.2026.01.003. URL http://dx.doi.org/10.1016/j.cell.2026.01.003.

Anthony J. Quartararo, Zachary P. Gates, Bente A. Somsen, Nina Hartrampf, Xiyun Ye, Arisa Shimada, Yasuhiro Kajihara, Christian Ottmann, and Bradley L. Pentelute. Ultra-large chemical libraries for the discovery of high-affinity peptide binders. *Nature Communications*, 11 (1):3183, Jun 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-16920-3. URL https://doi.org/10.1038/s41467-020-16920-3.

Stephen A. Rettie, Katelyn V. Campbell, Asim K. Bera, Alex Kang, Simon Kozlov, Yensi Flores Bueso, Joshmyn De La Cruz, Maggie Ahlrichs, Suna Cheng, Stacey R. Gerben, Mila Lamb, Analisa Murray, Victor Adebomi, Guangfeng Zhou, Frank DiMaio, Sergey Ovchinnikov, and Gaurav Bhardwaj. Cyclic peptide structure prediction and design using AlphaFold2. *Nature Communications*, 16(1):4730, May 2025a. ISSN 2041-1723. doi: 10.1038/s41467-025-59940-7. URL https://doi.org/10.1038/s41467-025-59940-7.

Stephen A. Rettie, David Juergens, Victor Adebomi, Yensi Flores Bueso, Qinqin Zhao, Alexandria N. Leveille, Andi Liu, Asim K. Bera, Joana A. Wilms, Alina Üffing, Alex Kang, Evans Brackenbrough, Mila Lamb, Stacey R. Gerben, Analisa Murray, Paul M. Levine, Maika Schneider, Vibha Vasireddy, Sergey Ovchinnikov, Oliver H. Weiergräber, Dieter Willbold, Joshua A. Kritzer, Joseph D. Mougous, David Baker, Frank DiMaio, and Gaurav Bhardwaj. Accurate de novo design of high-affinity protein-binding macrocycles using deep learning. *Nature Chemical Biology*, 21(12):1948–1956, 2025b. doi: 10.1038/s41589-025-01929-w. URL https://doi.org/10.1038/s41589-025-01929-w.

Joseph M. Rogers, Toby Passioura, and Hiroaki Suga. Nonproteinogenic deep mutational scanning of linear and cyclic peptides. *Proceedings of the National Academy of Sciences*, 115 (43):10959–10964, October 2018. ISSN 1091-6490. doi: 10.1073/pnas.1809901115. URL http://dx.doi.org/10.1073/pnas.1809901115.

James P. Roney and Sergey Ovchinnikov. State-of-the-art estimation of protein model accuracy using alphafold. *Phys. Rev. Lett.*, 129:238101, Nov 2022. doi: 10.1103/PhysRevLett.129.238101. URL https://link.aps.org/doi/10.1103/PhysRevLett.129.238101.

James P. Roney, Chenxi Ou, and Sergey Ovchinnikov. Protein Diffusion Models as Statistical Potentials. *bioRxiv*, 2025. doi: 10.64898/2025.12.09.693073. Preprint.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen Mckeown, and Rajesh Ranganath. A General Framework for Inference-time Scaling and Steering of Diffusion Models. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 55810–55827. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/singhal25b.html.

Nicholas J. Skelton, Michael F.T. Koehler, Kerry Zobel, Wai Lee Wong, Sherry Yeh, M. Theresa Pisabarro, Jian Ping Yin, Laurence A. Lasky, and Sachdev S. Sidhu. Origins of pdz domain ligand specificity. *Journal of Biological Chemistry*, 278(9):7645–7654, February 2003. ISSN 0021-9258. doi: 10.1074/jbc.m209751200. URL http://dx.doi.org/10.1074/jbc.M209751200.

Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy John O'Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, Mateo Reveiz, Roman Bushuiev, Tomáš Pluskal, Josef Sivic, Karsten Kreis, Arash Vahdat, Shamayeeta Ray, Jonathan T. Goldstein, Andrew Savinov, Jacob A. Hambalek, Anshika Gupta, Diego A. Taquiri-Diaz, Yaotian Zhang, A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, Chang C. Liu, Gene-Wei Li, Denes Hnisz, David M. Sabatini, William F. DeGrado, Jeremy Wohlwend, Gabriele Corso, Regina Barzilay, and Tommi Jaakkola. Boltzgen: Toward universal binder design. *bioRxiv*, 2025. doi: 10.1101/2025.11.20.689494.

Protenix Team, Milong Ren, Jinyuan Sun, Jiaqi Guan, Cong Liu, Chengyue Gong, Yuzhe Wang, Lan Wang, Qixu Cai, Wenzhi Ma, Yuxuan Zhang, Zhenyu Liu, Hanyu Zhang, Xinshi Chen, and Wenzhi Xiao. PXDesign: Fast, Modular, and Accurate De Novo Design of Protein Binders. *bioRxiv*, 2025. doi: 10.1101/2025.08.15.670450. URL https://doi.org/10.1101/2025.08.15.670450. Preprint.

Fayçal Touti, Zachary P. Gates, Anupam Bandyopadhyay, Guillaume Lautrette, and Bradley L. Pentelute. In-solution enrichment identifies peptide inhibitors of protein–protein interactions. *Nature Chemical Biology*, 15(4):410–418, March 2019. ISSN 1552-4469. doi: 10.1038/s41589-019-0245-2. URL http://dx.doi.org/10.1038/s41589-019-0245-2.

Quentin Trolliet, Alex Abrudan, Aryan Bhasin, Yunguan Fu, Joao Paulo Euko, Cheng Zhang, and Francesco Saccon. Chemical Descriptors and Deep Learning Embeddings for Scoring de novo Peptide Designs. *bioRxiv*, 2026. doi: 10.64898/2026.09.10.750670. URL https://doi.org/10.64898/2026.09.10.750670. Preprint.

Fanhao Wang, Yuzhe Wang, Laiyi Feng, Changsheng Zhang, and Luhua Lai. Target-Specific De Novo Peptide Binder Design with DiffPepBuilder. *Journal of Chemical Information and Modeling*, 64(24):9135–9149, 2024. doi: 10.1021/acs.jcim.4c00975.

Yuzhe Wang, Fanhao Wang, Laiyi Feng, Changsheng Zhang, and Luhua Lai. DiffPepDock : Efficient protein–peptide docking and binder screening via SE (3)-equivariant diffusion. *Protein Science*, 34(11):e70338, 2025a. doi: 10.1002/pro.70338. URL https://doi.org/10.1002/pro.70338.

Zhe Wang, Ruizhi Lai, Xinpei Wang, Xu Chen, Youjian Zhou, Shengbin Li, Xiaohui Qiu, Zekai Zeng, Jianye Yuan, Jinghuan Mao, Zhidong Chen, and Junqing Wang. Targeted Penetrating Motif Engineering of BH3 Mimetic: Harnessing Non-Canonical Amino Acids for Coinhibition of MCL-1 and BCL-xL in Acute Myeloid Leukemia. *Advanced Science*, 12(27):2503682, April 2025b. ISSN 2198-3844. doi: 10.1002/advs.202503682. URL http://dx.doi.org/10.1002/advs.202503682.

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, 2023. doi: 10.1038/s41586-023-06415-8. URL https://doi.org/10.1038/s41586-023-06415-8.

Ziyi Yang, Hanyuan Xie, Yinjun Jia, Xiangzhe Kong, Jiqing Zheng, Ziting Zhang, Yang Liu, Lei Liu, and Yanyan Lan. Cpsea: Large-scale cyclic peptide-protein complex dataset for machine learning in cyclic peptide design. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference. Curran Associates, Inc., 2025. doi: 10.52202/085713-1479. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/3eec719ab86712d32b065c5977f94ad0-Paper-Datasets_and_Benchmarks_Track.pdf.

Vinicius Zambaldi, David La, Alexander E. Chu, Harshnira Patani, Amy E. Danson, Tristan O. C. Kwan, Thomas Frerix, Rosalia G. Schneider, David Saxton, Ashok Thillaisundaram, Zachary Wu, Isabel Moraes, Oskar Lange, Eliseo Papa, Gabriella Stanton, Victor Martin, Sukhdeep Singh,

Lai H. Wong, Russ Bates, Simon A. Kohl, Josh Abramson, Andrew W. Senior, Yilmaz Alguel, Mary Y. Wu, Irene M. Aspalter, Katie Bentley, David L. V. Bauer, Peter Cherepanov, Demis Hassabis, Pushmeet Kohli, Rob Fergus, and Jue Wang. De novo design of high-affinity protein binders with AlphaProteo. *arXiv preprint arXiv:2409.08022*, 2024. doi: 10.48550/arXiv.2409.08022.

Silong Zhai, Huifeng Zhao, Jike Wang, Shaolong Lin, Tiantao Liu, Shukai Gu, Dejun Jiang, Huanxiang Liu, Yu Kang, Xiaojun Yao, and Tingjun Hou. PepPCBench is a Comprehensive Benchmarking Framework for Protein–Peptide Complex Structure Prediction. *Journal of Chemical Information and Modeling*, 65(16):8497–8513, 2025. doi: 10.1021/acs.jcim.5c01084. URL https://doi.org/10.1021/acs.jcim.5c01084.

Chenhao Zhang, Chengyun Zhang, Tianfeng Shang, Ning Zhu, Xinyi Wu, and Hongliang Duan. Highfold: accurately predicting structures of cyclic peptides and complexes with head-to-tail and disulfide bridge constraints. *Briefings in Bioinformatics*, 25(3):bbae215, 2024. doi: 10.1093/bib/bbae215.

Qiaoxuan Zhang, Yifei Geng, Haitao Chen, Jiaping Ni, Lixia Guan, Dong qing Zhai, Yuting Wang, Shengtao Xu, Miao-Miao Niu, Lufeng Zheng, and Weiwei Hu. Discovery of a Novel Dual-Targeting KRASG12D/HDAC Peptide Inhibitor for the Treatment of Pancreatic Cancer. *Journal of Medicinal Chemistry*, 69(5):5425–5440, February 2026a. ISSN 1520-4804. doi: 10.1021/acs.jmedchem.5c02435. URL http://dx.doi.org/10.1021/acs.jmedchem.5c02435.

Yuxuan Zhang, Chengyue Gong, Jinyuan Sun, Jiaqi Guan, Milong Ren, Song Xue, Hanyu Zhang, Wenzhi Ma, Zhenyu Liu, Xinshi Chen, and Wenzhi Xiao. Protenix-v2: Broadening the Reach of Structure Prediction and Biomolecular Design. *bioRxiv*, 2026b. doi: 10.64898/2026.04.10.717613. URL https://doi.org/10.64898/2026.04.10.717613. Preprint.

Yuxuan Zhang, Chengyue Gong, Hanyu Zhang, Wenzhi Ma, Zhenyu Liu, Xinshi Chen, Jiaqi Guan, Lan Wang, Yanping Yang, Yu Xia, and Wenzhi Xiao. Protenix-v1: Toward high-accuracy open-source biomolecular structure prediction. *bioRxiv*, 2026c. doi: 10.64898/2026.02.05.703733. URL https://www.biorxiv.org/content/early/2026/02/22/2026.02.05.703733.1.

Ning Zhu, Yanyu Ming, Chengyun Zhang, Sen Cao, Chongyang Li, Yaling Wu, Jingjing Guo, and Hongliang Duan. A data resource of peptide–protein interaction structures and binding affinities for peptide drug discovery. *Scientific Data*, Jul 2026. ISSN 2052-4463. doi: 10.1038/s41597-026-07962-1. URL https://doi.org/10.1038/s41597-026-07962-1.

APPENDIX

A Protein-Peptide Complex Structure Prediction **19**

A.1 Benchmark settings 19

A.2 Implementation details 21

A.3 Full benchmark results 23

B Peptide Binding Prediction **29**

B.1 Datasets and evaluation protocol 29

B.2 Full benchmark results 30

C Global Proposal **33**

C.1 Method details 33

C.2 Infrastructure optimizations 34

C.3 Benchmark and baseline settings 34

C.4 Additional benchmark results 37

C.5 Additional visualizations 42

D Local Refinement **44**

D.1 Method details 44

D.2 Benchmark results 45

E Designing Multi-target Binders and Macrocyclic Glues **48**

E.1 Method details 48

E.2 Additional visualizations 50

A Protein-Peptide Complex Structure Prediction

A.1 Benchmark settings

Dataset curation. We combined target-sequence clustering at 40% identity with manual inspection of deposited complexes to assemble a chemically diverse panel. Inspection verified that each peptide is a standalone ligand bound to its biological target, excluding fragments or termini excised from a larger protein. We restricted peptide lengths to 5–30 residues and selected peptides composed of the 20 canonical amino acids and target inputs containing at most 750 residues. Cyclization assignments required deposited covalent connectivity or an explicit publication/deposition annotation supported by the native endpoint geometry. The final panel contains eight head-to-tail, eight disulfide, and ten isopeptide complexes (Table 1). The isopeptide subset comprises five sidechain-to-sidechain lactams, three head-to-sidechain lariats, and two sidechain-to-tail lariats.

Table 1: Protein–peptide complexes used for the structural benchmark. Endpoint numbers use one-based positions in the modeled peptide; disulfide entries list SG–SG residue pairs, and other atom names follow the deposited residue nomenclature.

PDB	Target	Length	Topology	Bond endpoints
7K2F	KEAP1	7	Head-to-tail	N1–C7
5XN3	SPSB2	8	Head-to-tail	N1–C8
9CDZ	MDM2	16	Head-to-tail	N1–C16
9HGD	GABARAP	13	Head-to-tail	N1–C13
3AV9	HIV-1 integrase	8	Head-to-tail	N1–C8
6N7Q	AMA1	13	Head-to-tail	N1–C13
9CDT	MCL1	16	Head-to-tail	N1–C16
9CDU	RbtA	14	Head-to-tail	N1–C14
1VPP	VEGFA	20	Disulfide	7–15
3P72	GP1BA	11	Disulfide	1–11
4IB5	CSNK2A1	13	Disulfide	2–12
4W50	EPHA4	12	Disulfide	4–12
5VLL	PCSK9	14	Disulfide	1–14
8ZRT	EDNRB	21	Disulfide	1–15; 3–11
9GRM	CDC42	16	Disulfide	4–13
9MAP	PD-L1	28	Disulfide	2–28; 6–24; 10–16
2YQ7	BCL-XL	18	Sidechain lactam	CD8–NZ12
5LSO	SPF45	6	Sidechain lactam	NZ1–CD6
5WGD	ESR1	11	Sidechain lactam	NZ6–CG10
8ACK	PaAP	9	Sidechain lactam	CD1–NZ9
6SBA	TEAD4	20	Sidechain lactam	CD3–NZ18
1E4W	TAB2 Fab	7	Head lariat	N1–CD7
5TXE	AtxE2	18	Head lariat	N1–CG9
9KDF	EDNRB	15	Head lariat	N1–CG9
1URC	CCNA2	5	Tail lariat	NZ2–C5
2WMA	CCNA2	5	Tail lariat	NZ2–C5

Structure preparation. We retained the relevant biological assembly and mapped the peptide consistently between the deposited structure and model input. Target templates contain only observed target coordinates; missing target segments remain in the sequence without fabricated template coordinates. For 7K2F we used biological assembly 2. For 2WMA we retained the peptide-contacting CCNA2 chain and excluded the noncontacting CDK2 chain. For 9KDF we reduced the deposited fusion construct to the 387-residue EDNRB input, retaining all 33 peptide-contacting target residues and leaving the restored intracellular-loop segment untemplated. Noninteracting terminal caps in 1URC, 2YQ7, 5WGD and 6SBA were omitted; these caps do not form the evaluated cyclization bond. The noninteracting target palmitoyl modification in 6SBA was normalized to cysteine.

Prediction and evaluation. We evaluated AF3 (Abramson et al., 2024), Protenix v1 (`protenix_base_default_v1.0.0`) (Zhang et al., 2026c), and Protenix v2 (`protenix-v2`)

(Zhang et al., 2026b) using target-only templates and no MSA. Paired runs use seeds 2026–2028 and five generated diffusion samples per seed; the highest-ranked sample under each model's ranking score is retained for the displayed comparisons. Global metrics comprise target-aligned peptide-backbone RMSD, peptide-interface backbone RMSD, native-contact F1, and DockQ, computed with DockQ v2.1.3 (Mirabello & Wallner, 2024). For symmetric target dimers in 3AV9 and 1VPP, we evaluate equivalent chain assignments and reuse the assignment with the lowest target-aligned peptide-backbone RMSD for all pose metrics.

Local geometry is assessed through endpoint distances, bond angles, and disulfide torsion or amide nonplanarity. Reported distance-closure counts use a tolerance of 0.20 Å around the prediction's ordinary C–N median for amides and 0.30 Å around 2.03 Å for disulfides. Within each setting and seed, multiple bonds are first reduced to a case median; points then show the median across seeds and whiskers their range. Summary diamonds give equal weight to each case. Reference bands span the 5th–95th percentiles of deposited disulfide distances or ordinary amide geometry in the default predictions. Thus each figure presents both binding-pose recovery and the geometry of the intended linkage.

Inspection of existing cyclic-peptide benchmarks. HighFold encodes head-to-tail and disulfide shortcuts through its cyclic position-offset matrix (Zhang et al., 2024). Its two complex evaluation sets contain 17 structures each, covering five targets in the ADCP-derived set and seven in the external set, with ten distinct targets overall. Twelve ADCP complexes bind HIV-1 integrase; five external complexes bind trypsin and five bind plasmin. Together, integrase/LEDGF-derived and SFTI-related series account for 25 of the 34 complexes. The grouped native structures show repeated scaffolds and closely related binding poses (Fig. 8). Several SFTI-derived examples combine head-to-tail closure with disulfides, while some carry additional modifications. We therefore separate pure head-to-tail, disulfide, and isopeptide classes in the present benchmark to resolve their individual representation requirements.

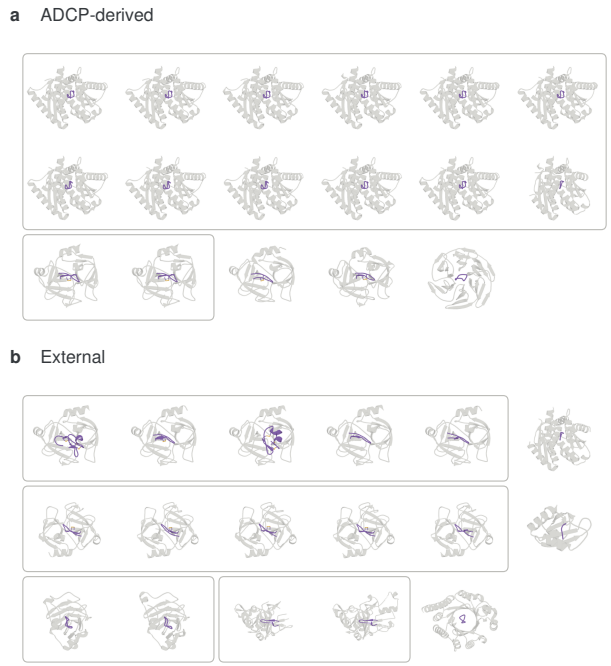


Figure 8: Native complexes in the HighFold evaluation sets. a, The 17 ADCP-derived complexes. b, The 17 external complexes. Rounded groups collect repeated biological targets. Peptides are violet, targets grey, and peptide disulfides gold. Views share a camera and scale within each target group; molecular sizes are not directly comparable across different target groups. Non-disulfide chemical modifications are not highlighted by this cartoon representation.

A.2 IMPLEMENTATION DETAILS

Chemical connectivity and model features. We treat the molecular bond graph, relative positional encoding, and token-bond features as separate inputs. Our AF3 adaptation preserves declared polymer–polymer links through structure cleaning, atom featurization and output reconstruction. Head-to-tail and crosslink endpoints are supplied through `bondedAtomPairs`; Protenix inputs use `is_cyclic` and explicit covalent-bond declarations. Product-forming preprocessing removes the leaving atoms associated with the declared amide, including terminal OXT for head-to-tail closure and the appropriate carboxyl oxygen for an isopeptide linkage. These changes retain the pretrained weights and feature dimensions.

Cyclic RPE replaces linear residue offsets within the peptide by signed shortest-path offsets around its head-to-tail backbone. The all-edge variant additionally includes declared crosslinks as graph shortcuts. Offsets are computed before the existing distance clipping and embedding; the surrounding target representation is retained. A token-bond feature instead marks the connected residue pair in the model's binary pairwise bond channel. It conveys connectivity without a chemistry-specific target bond length. Accordingly, the baseline feature setting uses native offsets and no additional token hint while preserving the specified chemical graph.

For the isopeptide feature comparison, we hold an open chemical state fixed while varying RPE and bond hints. Additional matched controls compare open and linked-product atom inventories and canonical versus neutralized endpoint charges. These controls reproduce the main topology-dependent trends: bond hints promote endpoint approach, most clearly in head lariats, while their effects on pose recovery depend on the complex. The cross-model isopeptide comparison uses the exact linked-product inventory. This separates the feature response from changes in the underlying chemical construction.

Repairing the Protenix v1 head-to-tail seam. With a linear atom layout, residues joined across the cyclic seam can lie outside one another's local attention support. Protenix v1 consequently retained stretched terminal C–N distances even with the correct atom graph, leaving-atom removal and cyclic RPE (Fig. 9). We addressed this by deterministically reordering whole peptide residues to reduce the atom-layout separation of connected residues. Atom features and diffusion coordinates are gathered consistently into this internal order, while token, MSA and template residue order remains fixed. Noise is generated in public atom identity before permutation, and predicted coordinates are restored to the public order at output. The existing 32-query/128-key local attention windows are retained.

In matched diagnostic runs, this permutation repaired the six failing MDM2/GABARAP case–seed combinations while preserving the short KEAP1 control. Removing the seam token hint alone did not repair the long-seam failures. The subsequent eight-case head-to-tail control closed 9/24 rank-0 seams with native Protenix v1 routing and 24/24 with the reordered cyclic-RPE setting. Protenix v2 closed all 24/24 seams with native routing, so the reorder operation remains an optional, checkpoint-specific adaptation.

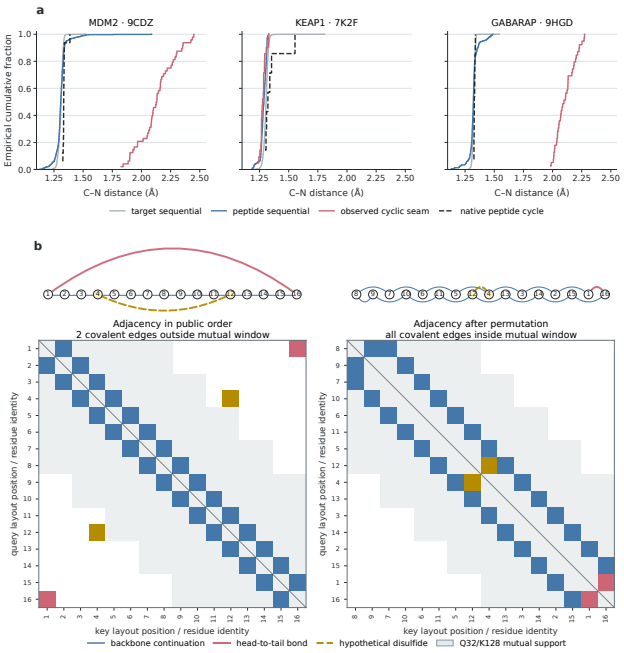


Figure 9: Head-to-tail seam geometry in Protenix v1. a, Empirical cumulative distributions from all 108 retained Attempt 1 predictions. b, Residue-order permutation, adjacency, and local-attention-window schematics. The yellow disulfide is hypothetical; the attention illustration assumes eight atoms per residue and does not measure atom-level attention in the predicted complexes.

A.3 FULL BENCHMARK RESULTS

Head-to-tail representations. The complete AF3 comparison in Fig. 10 evaluates native features, cyclic RPE, token bonds, and their combination. Cyclic RPE recovers the terminal C–N linkage across all eight cases and three seeds, closing 24/24 selected seams. Token bonds alone close 11/24, and the combined setting closes 22/24. Cyclic RPE also improves global pose recovery across much of the panel. These results favor encoding cyclic backbone distances directly.

Disulfide representations. Native features recover near-native S–S geometry across most cases. Adding all-edge RPE produces elongated or otherwise distorted disulfides and worsens pose recovery for several complexes. Token-bond features instead compress the S–S linkage: all 33 retained token-only bond observations are shorter than their corresponding deposited bonds, with the plotted case-weighted median at 1.60 Å. The combined setting also distorts local geometry. The same compression tendency is observed in Protenix v1 feature comparisons. We retain the disulfide chemical graph while leaving its RPE and token-bond hints off.

Isopeptide representations. The ten-case panel spans sidechain lactams and both lariat orientations. Native features already recover several binding poses accurately. Token-bond hints bring the intended amide endpoints closer, especially for head lariats, but improvements in endpoint distance do not consistently track DockQ. All-edge RPE is less effective across this heterogeneous set. The matched chemistry and charge controls preserve these qualitative trends. We therefore use token-bond hints to guide isopeptide endpoint approach and assess the resulting linkage geometry explicitly during design.

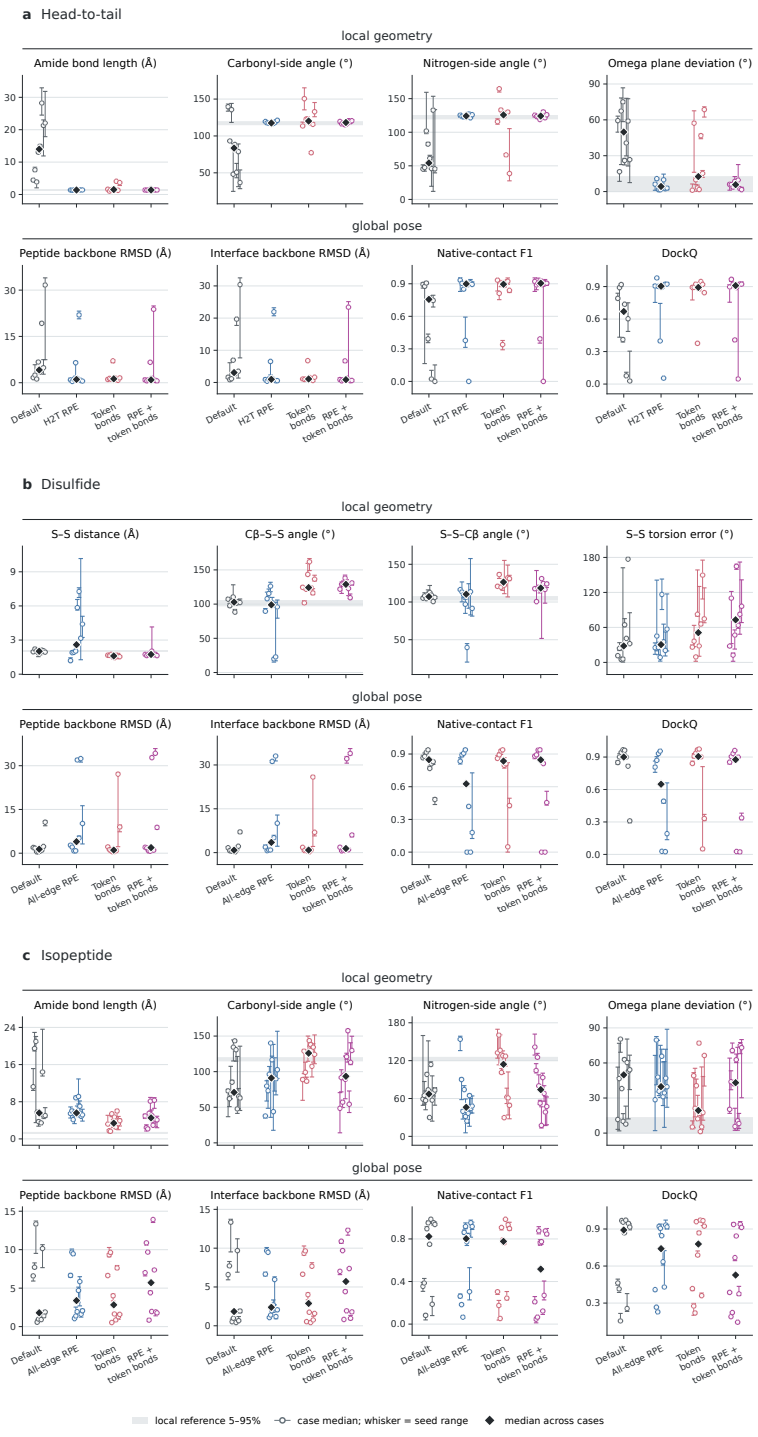


Figure 10: Complete structural benchmarks for eight head-to-tail, eight disulfide, and ten isopeptide cases. Rank-0 predictions compare native features, topology RPE, token bonds, and their combination. Local panels show bond distance, angles and torsion or nonplanarity; global panels show target-aligned peptide and interface RMSD, native-contact F1 and DockQ. The isopeptide feature ablation holds the open chemical state fixed. Bond measurements are reduced within case, setting and seed; points and whiskers show case medians and seed ranges. Black diamonds weight cases equally. Reference bands retain the definitions in the benchmark protocol.

Robustness to cyclic sequence origin. A cyclic peptide can be serialized from any residue. We compared Protenix v1 predictions across cyclic rotations while holding the physical sequence, target, chemical graph and canonical atom-identity noise mapping fixed. Both arms use residue reordering and differ in whether the cyclic relation is encoded by RPE or a seam token-bond hint. The displayed four-case comparison contains 132 rank-0 predictions per representation across three seeds (Fig. 11). Cyclic RPE reduces sensitivity to the chosen origin and improves seam geometry. The case-level curves also expose residual origin dependence, including the short KEAP1 peptide. The resulting choice of cyclic RPE prioritizes robustness without adding a separate sequence-origin search.

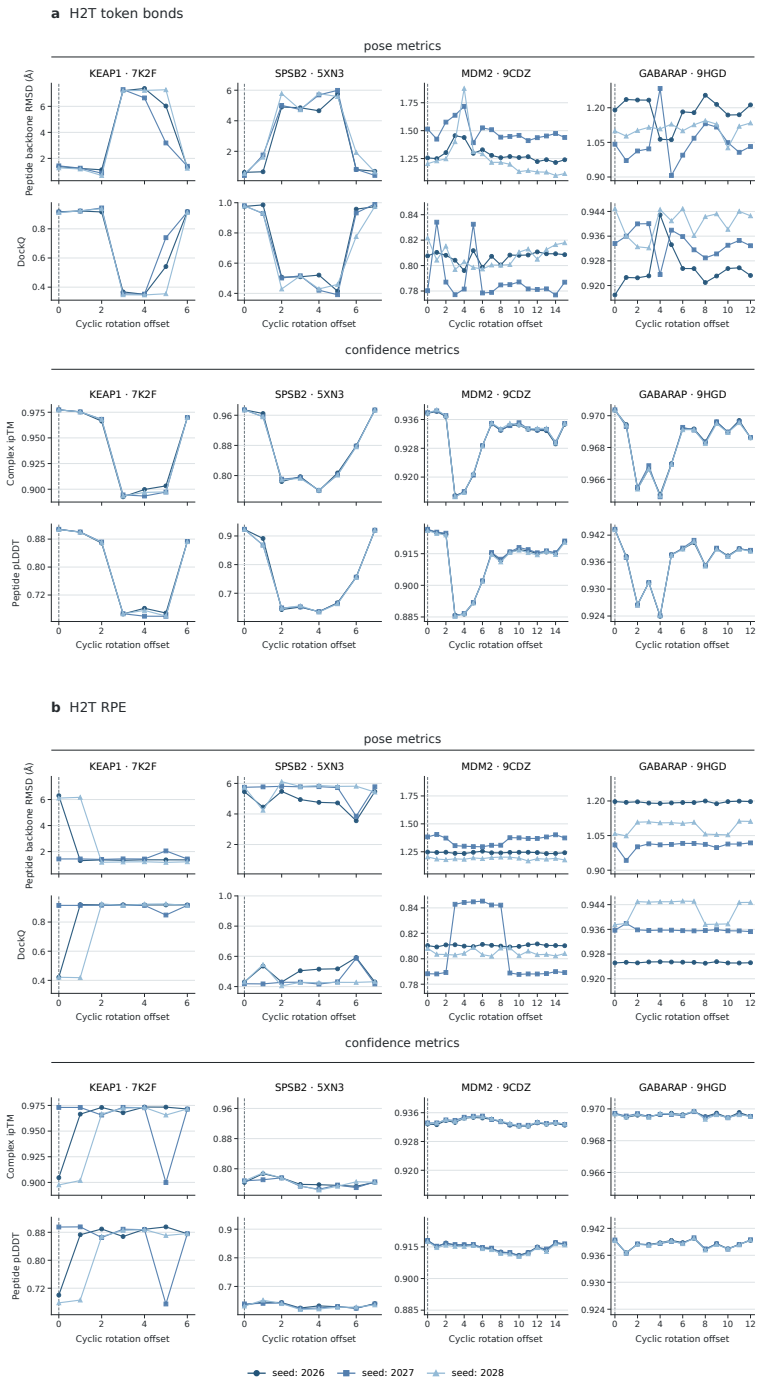


Figure 11: Robustness to sequence rotation across four head-to-tail cases and three seeds. Each representation retains 132 rank-0 predictions, selected from five samples per job. Both Protenix v1 arms use residue reordering; only the cyclic relational feature differs. Panels use matched axis limits across representations and preserve individual seed results. Peptide atoms are mapped back to their physical identities before comparison with the native structure.

Comparison of AF3 and Protenix checkpoints. We compare the same cases and seeds using topology-specific settings (Fig. 12). Head-to-tail inputs use cyclic RPE, with residue reordering in the selected Protenix v1/v2 arms. Disulfide inputs use native offsets and no token hints, again with reordering in the selected Protenix arms. Isopeptide inputs use linked-product chemistry without additional RPE or token hints, with reordering for v1 and native routing for AF3/v2.

All three models recover head-to-tail structures well and close all 24 selected seams. Protenix v2 improves on v1 for pure disulfides; AF3, v1 and v2 respectively recover 30/33, 19/33 and 25/33 S–S distances within the benchmark closure criterion. Isopeptide pose recovery is strongest overall for AF3, while covalent closure remains difficult across the models. These comparisons support AF3 as the global-proposal predictor and the topology-specific feature choices described above.

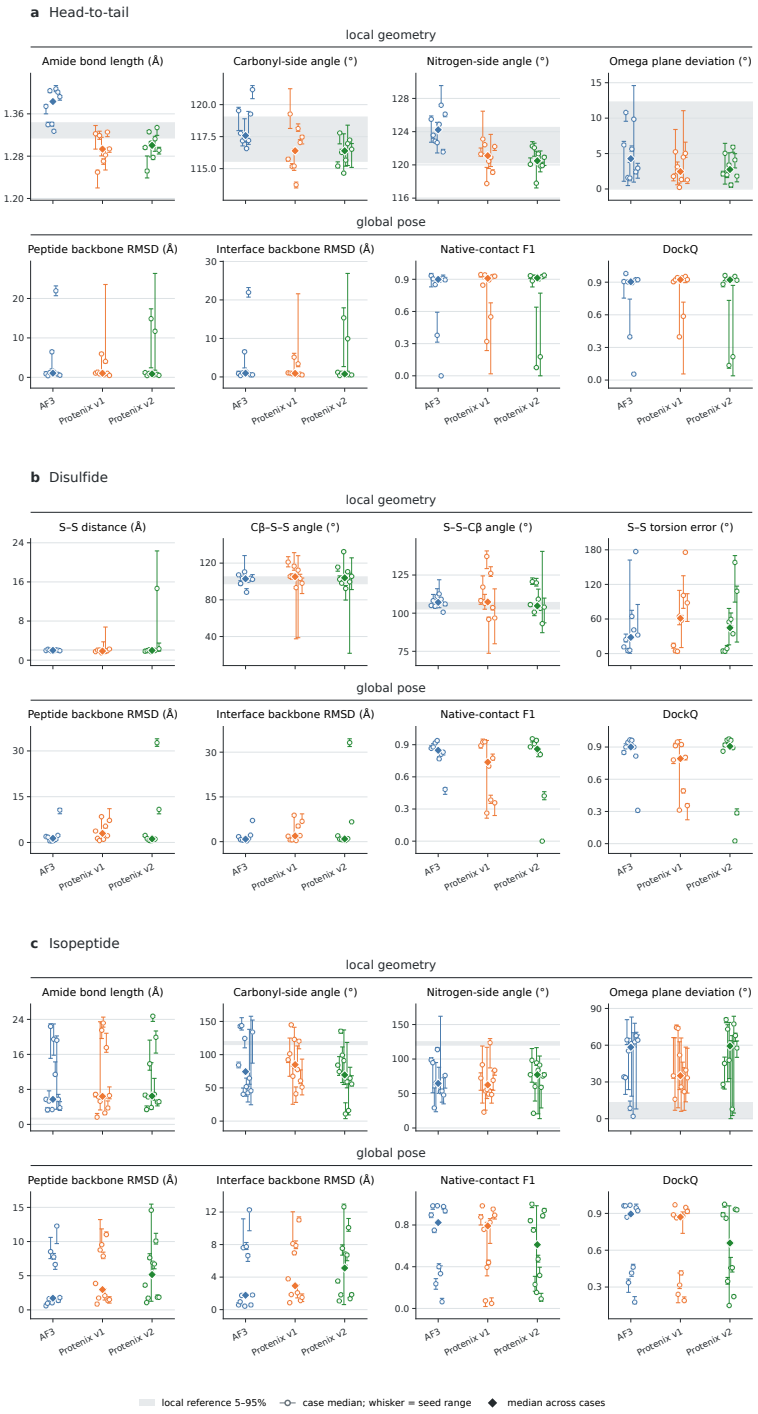


Figure 12: Cross-model comparison of cyclic peptide structure prediction for eight head-to-tail, eight disulfide and ten isopeptide complexes. Each model uses three paired seeds and its highest-ranked sample per seed. Head-to-tail arms use cyclic RPE; disulfide arms use native RPE without token hints; isopeptide arms use linked-product chemistry without additional RPE or token hints. Local geometry and global pose panels retain the case-weighted reductions and reference bands of the structural benchmark.

B PEPTIDE BINDING PREDICTION

B.1 DATASETS AND EVALUATION PROTOCOL

Database-derived peptide affinities. We constructed the PPIKB cohort from the database's main affinity export, containing 21,845 records at retrieval (Zhu et al., 2026). We selected canonical L-amino-acid/Gly peptides of 8–18 residues with explicit, supported terminal chemistry and cyclization connectivity, a single target UniProt assignment, and positive quantitative K_D values. Source endpoint and unit fields were checked for consistency. Linear peptides, disulfides, head-to-tail rings, and resolvable lactam/lariat linkages were eligible. We excluded the publication flagged for affinity transcription errors by Trolliet et al. (2026) and retained original measurement identifiers throughout processing. This screen yielded 3,361 records corresponding to 2,353 target–peptide identities across 741 source UniProt groups.

We selected source targets with at least 15 peptide identities and excluded the two MHC groups. The resulting 21 source-target groups were resolved into 30 modeling contexts, separating binding domains, known target variants, and corrected biological target identities. For example, HBV variants were modeled separately, and source records referring to avidin or hoefavidin were separated from streptavidin. We preferred deposited target structures and used AlphaFold Database domain templates when the selected domain had mean Cα pLDDT above 80. Supported multichain binding units were retained. Target structures supplied templates without peptide coordinates.

Terminal caps were retained in measurement identity but omitted from the modeled peptide. Records sharing the same bare sequence, topology and target context were assigned one prediction. When capped and uncapped measurements mapped to that model, uncapped records took precedence; otherwise, the mapped capped records supplied the reference endpoint. We aggregated selected repeat measurements by median pK_D, with $\mathrm{p}K_D = 9 - \log_{10}(K_D\,[\mathrm{nM}])$. Of 426 distinct model inputs, seven were excluded because of topology-gate failure, failed Rosetta evaluation, or invalid separated-chain energy outputs. The final figure reapplies the minimum of 15 retained models per target context, giving 368 models across 17 contexts. Predictor confidence was recorded without an additional confidence-based exclusion.

Primary-literature mutation series. We used an automated, iterative literature-curation workflow in Codex with GPT-5.6 Sol and GPT-6 Astra. The workflow retrieved primary papers and supporting information, extracted sequence identities and binding measurements, checked assay and chemical context against the original tables, and linked eligible families to experimentally determined complexes. Structured records retain source locations, experimental molecular identities, measurement relations, and residue mappings. The nine retained series contain 125 distinct canonical sequences, with 104 exact measurements and 21 numerical bounds (Table 2).

Within each series, peptides have equal length and share an author-defined family and a comparable assay cohort. We included all eligible same-length variants in the inspected source panel, including redesigns with multiple substitutions. We required a quantitative direct-binding or competition endpoint and an experimentally supported pose reference for a measured member or an explicitly linked precursor/segment. Measurements from different endpoints or assay cohorts were kept separate. Numerical bounds retain their reported relation and are shown in the figures; exact-value correlations use only uncensored measurements.

Table 2: Primary sources for the peptide mutation benchmark. Counts include numerical bounds, with exact measurements in parentheses. SPR, surface plasmon resonance; BLI, biolayer interferometry; FA/FP, fluorescence anisotropy/polarization; ITC, isothermal titration calorimetry. Anchors include exact assayed members and author-linked precursors or segments.

Series	n (exact)	Length	Assay / endpoint	PDB anchor	Primary source
PUMA–MCL1	5 (5)	27	SPR / K_D	2ROC	Rogers et al. (2018)
MDM2–pDI	6 (6)	13	BLI / K_D	3JZR	Touti et al. (2019)
Aurora-A–TACC3	15 (5)	15	Competition FA / IC_{50}	5ODT	Gimenez et al. (2025)

Series	n (exact)	Length	Assay / endpoint	PDB anchor	Primary source
PCNA–p21	45 (45)	15	SPR / K_D	7KQ1	Horsfall et al. (2021)
eIF4E–EE02	16 (6)	11	Cap competition / apparent K_D	7EZW	Frosi et al. (2022)
CAL PDZ	11 (10)	10	Competition FP / K_i	4JOK	Amacher et al. (2014)
Erbin PDZ	15 (15)	5	Bead competition / IC_{50}	1N7T	Skelton et al. (2003)
GRB2 nSH3–SOS1	6 (6)	12	ITC / K_D	4GBQ	McDonald et al. (2009)
BCL-xL–Bim	6 (6)	23	FA / K_D	3FDL	Foight et al. (2014)

Source-specific choices preserve assay comparability. The Aurora panel uses the reduced 522–536 TACC3 series, with no disulfide declaration. The eIF4E panel uses the cap-site competition endpoint and retains its physical Cys2–Cys10 disulfide. PCNA uses 45 measured 15-mers, including scaffold variants, and the full directly contacting PCNA chain as the target. CAL includes the 11 measurements from the selected study. For MDM2, main-figure values define the preferred endpoint; alternative supplementary fits remain separate. Models omit terminal caps and assay reporters while retaining the source-defined peptide sequence and physical cyclization bonds.

Prediction and scoring. Both cohorts use AF3 with target-only templates, empty MSAs, seed 2026, ten recycling cycles and five diffusion samples per candidate. We select the highest native-ranking sample without consulting measured binding, then perform one full-complex Rosetta relaxation with the `beta_jan25` score function and a maximum of 200 minimizer iterations. Confidence metrics are taken from the selected AF3 prediction before relaxation. Rosetta dG is the relaxed-complex interface-energy difference in Rosetta energy units (REU); negative dG is used so that larger values correspond to more favorable interactions.

We report peptide ipTM and $ipSAE_{min}$. Following the ipSAE construction (Dunbrack, 2025), partner residues are selected with interchain PAE below 10 Å, and the TM-like normalization depends on the number of qualifying partner residues. We compute both target-aligned and peptide-aligned directions and retain their minimum. All physical target chains are pooled as one partner group, excluding target–target entries. This definition is held fixed across both cohorts.

Pose basins and statistical summaries. Mutation-series basin membership uses peptide N/CA/C/O RMSD strictly below 2.5 Å after fitting the experimentally resolved, mapped target Cα atoms. Peptides are not independently superimposed. The reference correspondence is fixed from the source structure: coverage is 27/27, 12/13, 15/15, 15/15, 11/11, 6/10, 5/5, 10/12 and 23/23 residues for the nine series in table order. CAL uses its declared binding core at positions 5–10; the deposited N-terminal region includes lattice contacts. Basin labels in all score panels use the selected AF3 pose before Rosetta relaxation.

Spearman correlations are calculated within each target context or assay series against pK_D or the negative logarithm of the molar K_D, K_i or IC_{50} endpoint. Positive correlations indicate agreement with stronger binding. Figure panels a and b show one equally weighted correlation per context or series; panel b includes all exact observations irrespective of basin membership. Candidate-level bootstrap intervals use 5,000 resamples. Primary series-level tests use exhaustive permutations for at most eight exact observations and 19,999 random permutations otherwise, with Benjamini–Hochberg adjustment across the 27 series–metric comparisons. Experimental bounds are excluded from these correlations and displayed at their thresholds with arrows.

B.2 Full benchmark results

Heterogeneous screening and mutation-series ranking. In the 17 PPIKB contexts, median Spearman correlations are 0.203 for ipTM, 0.202 for $ipSAE_{min}$ and 0.146 for negative Rosetta dG. Across the nine curated assay series, the corresponding medians rise to 0.371, 0.600 and 0.527. The individ-

ual series in Fig. 13 show where each metric carries useful binding information. $\text{ipSAE}_{\min}$ reaches $\rho = 0.90$ for PUMA–MCL1 and 0.89 for eIF4E–EE02, while negative Rosetta dG reaches 0.77 for BCL-xL–Bim. MDM2 has negative correlations for all three metrics despite recovery of the reference pose in all six selected predictions.

PCNA supplies the largest series, with 45 exact measurements. Its correlations are 0.66, 0.77 and 0.57 for ipTM, $\text{ipSAE}_{\min}$ and negative dG, respectively; their candidate-bootstrap 95% intervals are [0.47, 0.80], [0.63, 0.86] and [0.35, 0.73]. These three associations retain an adjusted $q = 0.00045$. Among the 38 predictions within the reference basin, correlations remain positive at 0.59, 0.75 and 0.47. The strong interface-confidence signals complement the evidence for ipSAE in experimental protein-binder screening (Overath et al., 2025).

Binding information in pose recovery. For Aurora-A–TACC3, the strongest measured variant has $\text{IC}_{50} = 80\,\mu\text{M}$ and is the sole selected prediction assigned to the reference basin. The ten variants reported at $\text{IC}_{50} \geq 500\,\mu\text{M}$ all lie outside the basin. For eIF4E–EE02, the 11 outside-basin predictions comprise nine variants with apparent $K_D > 20\,\mu\text{M}$ and two with values of 15.03 and 16.71 μM. The parent and several stronger variants recover the reference binding mode. These observations motivate preserving a structurally supported basin during sequence optimization.

Across the full mutation set, 93/125 selected AF3 predictions lie within their mapped reference basins, and 89 are inside after Rosetta relaxation. The reference-basin analysis supplies a structural criterion alongside confidence and interface energy: pose consistency supports candidate eligibility, while local energetic comparisons guide refinement among eligible sequences. This motivates the separation of global proposal and basin-constrained local refinement in OmniPepBuilder.

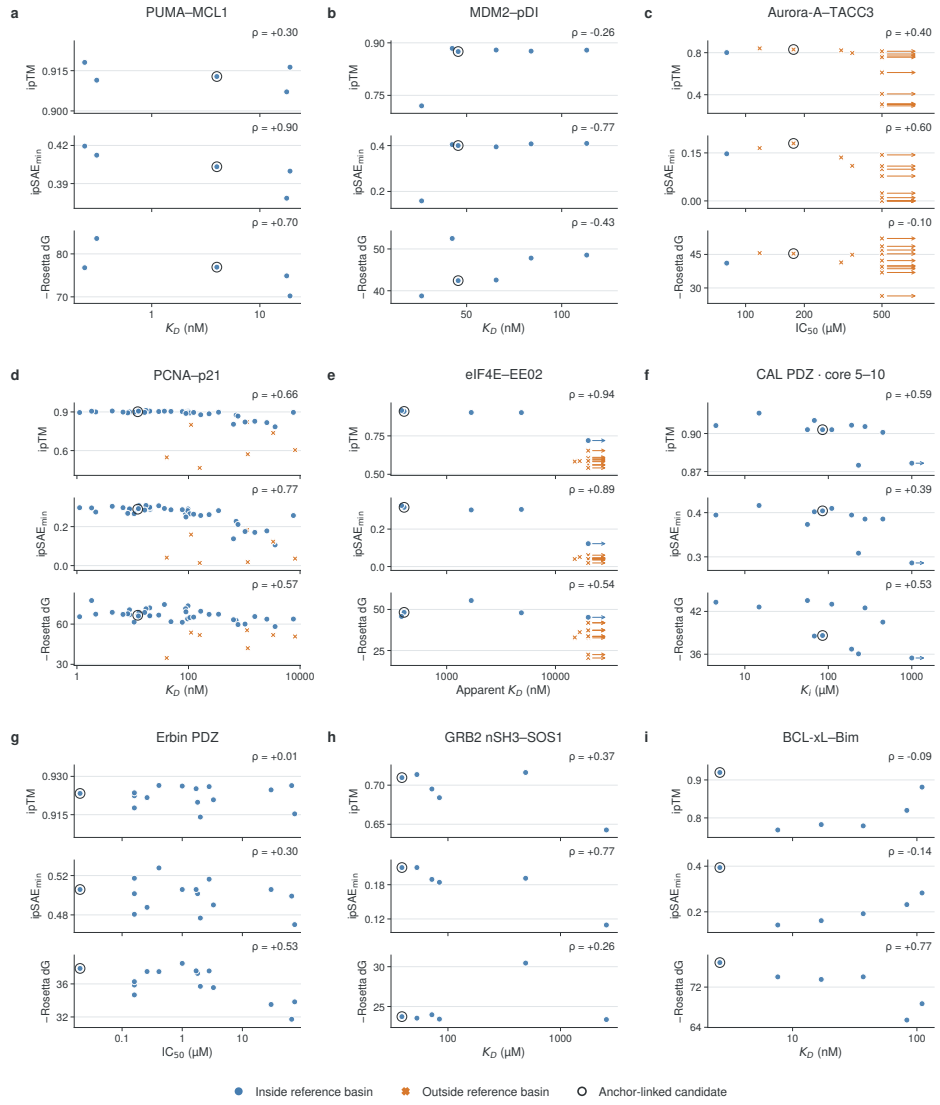


Figure 13: Prediction metrics versus measured binding for nine curated peptide series. Panels show peptide ipTM, ipSAE_min and negative Rosetta dG (REU). Experimental endpoints retain source units on logarithmic axes; larger values indicate weaker binding. Blue circles indicate mapped peptide-backbone RMSD below 2.5 Å after target alignment; orange crosses indicate larger RMSD. Basin membership uses the selected AF3 pose before relaxation. Black rings mark anchor-linked candidates. Mapped coverage for a–i is 27/27, 12/13, 15/15, 15/15, 11/11, 6/10, 5/5, 10/12 and 23/23 residues. Right arrows show bounds; all 21 bounds are excluded from correlations. Positive Spearman correlations indicate agreement with stronger binding. Caps and assay reporters were omitted from the models.

C GLOBAL PROPOSAL

C.1 METHOD DETAILS

Global proposal operates on sequence–structure states $s = (S, X, \tau)$, where τ records peptide length, fixed residue identities and resolved backbone/crosslink connectivity. Given the selected complex structure, ProteinMPNN proposes a sequence and AF3 predicts its complex with the target:

$$S' \sim q_\phi(S' \mid X, \tau), \qquad X' \sim p_\theta(X' \mid S', \tau, \text{target}).$$

These stochastic transitions can improve confidence, deteriorate or move to another pose. Each ordinary lineage follows one selected continuation per cycle. Fresh origins draw random canonical sequences without unconstrained cysteine; topology-defining residues remain fixed. A returned origin restores the exact selected structure and topology, generates a new ProteinMPNN child and starts a new lineage at cycle 1. The parent remains available for later returns, including after reaching its own cycle cap or being accepted. The cross-lineage restart edge preserves ancestry.

The AF3 profile uses model version 3.0.4, target-only coordinate templates, no target or binder MSA, ten recycles and five diffusion samples per candidate. ProteinMPNN uses the vanilla `v_48_020` checkpoint at sampling temperature 0.1. Every prediction sample is evaluated against the hard gates. For a single target assembly, the surviving sample with the highest native ranking score is selected, with a stable sample-ID tie break. Acceptance is then tested on this one representative: binder ipTM > 0.85 and peptide pLDDT > 0.80, in addition to the hotspot and topology gates. Each lineage ends at acceptance, hard-gate rejection, accepted-sequence duplication or eight prediction evaluations. The search permits eight active or reserved lineages; return probability is 0.70 when a selectable pool exists.

The pose history $\mathcal{V}$ contains target-mappable selected representatives from continuing, cycle-capped and accepted candidates. We define a direct similarity graph K on this history. Each prediction is aligned to a fixed target-reference patch comprising residues whose Cα atoms lie within 10 Å of any configured hotspot Cα. For permitted proper target symmetries, each correspondence is fitted to this same patch; the peptide receives the target transform without an independent peptide fit. The benchmark treats the equivalent sites of IL-17A and TNFα under their C2 and C3 correspondences, respectively. For allowed directed peptide registrations $\mathcal{M}_{su}$,

$$d(s,u) = \min_{g,h,\, m \in \mathcal{M}_{su}} \left[\frac{1}{4|m|} \sum_{(i,j)\in m} \sum_{a \in \{\mathrm{N},\mathrm{C}_\alpha,\mathrm{C},\mathrm{O}\}} \left\| T_{s,g} x^a_{s,i} - T_{u,h} x^a_{u,j} \right\|^2 \right]^{1/2}. \tag{4}$$

Here $T_{s,g}$ denotes the target-reference transform for symmetry correspondence g. Direct similarity requires $d(s,u) < 2.5$ Å and $\min(L_s, L_u)/\max(L_s, L_u) > 0.80$. Unequal lengths use contiguous shorter-to-longer windows. Head-to-tail backbones additionally allow directed cyclic and wrapped registrations; decorated linear backbones retain linear registration. Side-chain connectivity remains part of the exact restart state. The graph is symmetric, includes self-matches and need not be transitive.

To construct selectable support $\mathcal{S}$, we scan eligible states in descending confidence-tuple order with stable identifiers breaking ties. A state is suppressed only if an already retained, directly matching state has identical length, fixed residues and resolved topology, is no worse in both confidence coordinates, and is strictly better in at least one. Incomparable alternatives remain selectable. Recomputing this retained-witness rule from the history allows a suppressed state to reenter if its witness is no longer retained. No additional confidence threshold limits admission to this pool.

GROMOS (Daura et al., 1999) partitions $\mathcal{S}$ by repeatedly selecting the remaining state with the most remaining direct neighbors, assigning that state and its neighbors to a cell, and removing them from further assignment. Stable identifiers resolve ties. Every cell member remains a restart option. The partition is refreshed when retained support changes, permitting new cells and changed boundaries. Acceptance reporting applies the same clustering procedure to a separate set of accepted structures.

Eqs. 1 and 2 use each source's own confidence deficit. More generally, for configured metric j with scale $a_j > 0$, threshold θ_j and direction-aware deficit d_j, the implemented energy is $E_s = \max_j[d_j(s, \theta_j)/a_j]_+$. The benchmark uses unit scales for both confidence metrics, so a strong metric cannot offset the other metric's deficit. $E_s = 0$ denotes reaching the numerical thresholds; the strict acceptance comparisons and remaining gates are evaluated separately. E_s is dimensionless, and T controls allocation selectivity.

Return exposure is attached to exact historical sources. A pending return contributes one charge immediately, before another source is drawn. Materializing its first child retains that charge; later descendant evaluations add no return charge. An unmaterialized allocation marked failed releases its charge. Directly matching states share this history through H_s, including charges from sources that have since been suppressed. Repartitioning cells therefore does not reset exposure. The total weight of a cell is its mean exposure-adjusted source quality, $\sum_{s \in c} w_s = |c|^{-1} \sum_{s \in c} \exp(-E_s/T)/(1 + H_s)^{\alpha}$. This combines quality preference, density normalization and revisit pressure without restricting replay to a cell representative.

C.2 Infrastructure optimizations

Prediction uses one persistent model worker per allocated GPU. Compatible tasks reuse the loaded predictor, and each candidate attempt has an independent completion handle, allowing an idle device to receive work while other devices continue. A bounded CPU pool performs metric extraction and confidence compression concurrently with GPU prediction. Native ProteinMPNN tasks use available devices exclusively and release their processes on completion; the predictor remains resident. A single controller commits scientific state, updates the archive and allocates new origins. Ready predictions rotate across source/cycle groups, and pending origins count toward the active-lineage limit.

Archive geometry is updated incrementally. Transformed backbone coordinates, target symmetry correspondences and peptide registration indices are cached. A spatial index on ordered segment-centroid descriptors filters impossible matches using a lower bound on exact backbone RMSD; surviving pairs undergo the full symmetry- and registration-aware comparison. Each new/earlier pair is compared once, with direct matches stored as compact bitsets. Pareto retention and GROMOS then reuse the graph without rereading structures or recomputing RMSDs. The cell partition is reused when support is unchanged, while new return allocations update exposure before the next draw.

Bounded CPU measurements on retained pools of 3,420 linear-backbone and 3,657 cyclic-backbone states recorded initial graph construction times of 32.22 and 87.31 s, respectively. Adding the final 32 states took 0.151 and 0.865 s. Median cached snapshot times were 0.145 and 0.070 s, and allocation-distribution updates took 1.81 and 1.94 ms. These measurements include geometry and indexing but exclude structure-file parsing and archive writes. Thread counts were restricted to one. They characterize archive overhead rather than end-to-end design throughput.

Checkpointing preserves selected structures, topology, ancestry, direct matches, allocation records and the scheduler random-number state. Archive refreshes are coalesced at allocation/checkpoint boundaries. Task seeds and completion order are recorded, because asynchronous execution determines when evidence becomes available to the adaptive policy.

C.3 Benchmark and baseline settings

Benchmark settings. We evaluated eight targets spanning established peptide-binding grooves, receptor interfaces and more challenging binding sites, using the structures and hotspots in Table 3. Each target was paired with five peptide topologies: linear, head-to-tail, disulfide, Glu–Lys isopeptide and Asp–Lys isopeptide. For each new backbone or design trajectory, peptide length was sampled uniformly from the integers 8–18; subsequent sequence redesign preserved that length. Each method–case campaign was allocated 12 hours on one NVIDIA L40S, with repeated design attempts throughout the budget. Method coverage follows the implemented topology-specific workflows described below.

Table 3: Targets used for global-proposal benchmarking. Chain identifiers and residue numbers refer to the prepared source structures used by the benchmark configurations. Input lengths include unresolved residues retained in the query sequence. IL-17A and TNFα retain their dimeric and trimeric assemblies, respectively.

Target	Source PDB	Prepared target chains and residues	Input residues	Hotspots
MDM2	3G03	A:26–111	86	A:57, 61, 62, 67, 93, 100
MCL1	2PQK	A:172–321	150	A:231, 235, 253, 263, 267, 270
IL-23R	5MZV	C:24–126	103	C:28, 100, 108, 109, 113
PD-L1	8ALX	A:20–134	115	A:54, 56, 66, 113, 115, 123
KRAS G12D	5XCO	A:1–169	169	A:12, 61, 68, 69, 72, 95, 96, 99
CXCR4	3OE0	A:25–303	279	A:113, 117, 171, 187, 188, 190, 193, 262
IL-17A	5VB9	A, B:14–132 each	2×119	A:44, 53, 67; B:95
TNFα	5M2I	A, B, C:6–157 each	3×152	A:31, 32; C:87, 91

Target preparation preserved observed coordinates and chain boundaries. Missing regions remained in query sequences where supported but were omitted from coordinate templates; methods requiring fixed coordinate contigs received the observed segments. IL-17A uses the biological dimer, with the assembly-generated partner relabeled B. The CXCR4 input is a single-protomer derivative of 3OE0: fusion-distorted coordinates at residues 228–236 are omitted, S229/K230 are restored in sequence only, and W125/P240 are retained. KRAS G12D uses the nucleotide-free target derived from 5XCO. The TNFα benchmark uses the inter-protomer site listed in the table; the extended WP9 examples in Appendix C.5 use a separate hotspot set.

OmniPepBuilder and the no-replay reference. We used the AF3–ProteinMPNN implementation described in Appendix C.1 for all 40 target–topology cases. Both policies used the same prediction profile, sequence-design model, acceptance criteria and eight-cycle lineage cap. The no-replay reference initialized every lineage independently from a random sequence. Root and acceptance count limits were set to 10^6, leaving the time budget as the operative campaign limit.

RFdiffusion/RFpeptides. We used the code and released Complex checkpoint from the public RFdiffusion repository (Watson et al., 2023; Rettie et al., 2025b) for linear and head-to-tail peptides. Backbone generation used 50 denoising steps, native noise defaults and hotspot conditioning from Table 3; head-to-tail designs additionally used the RFpeptides cyclic encoding. ProteinMPNN assigned two sequences per linear backbone and four per cyclic backbone at temperature 0.1 and batch size one, with target chains fixed. We processed every completed backbone through sequence design and retained all resulting designs. This baseline used ProteinMPNN without FastRelax; common reprediction was performed separately. The separately measured sequence-design time was added to the backbone-generation budget.

BoltzGen. We used the code and released checkpoints from the public BoltzGen repository (Stark et al., 2025) for all five topologies. The native `peptide-anything` pipeline performed design, inverse folding, folding, analysis and filtering. Each invocation supplied ten independently sampled specifications, one design per specification and diffusion batch size one, split equally between the diverse and adherence design checkpoints. We retained the released inverse-folding and folding models and their native settings. Hotspot conditioning was enabled, and crosslink endpoint identities and intended bonds were included in the specifications. All final analysis-table entries entered the native pool, with their native filter flags retained.

AnewOmni. We used the code and released `model.ckpt` from the public AnewOmni repository (Kong et al., 2026a) for linear, head-to-tail and terminal-Cys disulfide peptides. Native peptide templates used trial size eight, graph guidance 1.0 and hotspot conditioning with a 10 Å pocket. Length bounds were set to `size_min=8` and `size_max=19`, following the implementation's exclusive upper bound. Generation and native physical-validity, confidence and chirality filters were retained. We saved every completed native filter survivor without the rolling top-50 retention limit. The disulfide template fixed cysteines at the two termini. Eight retained campaigns ended before 12 hours; their rates use the actual generation exposure.

HalluDesign. We used the public HalluDesign implementation (Fang et al., 2025), including its customized AF3 backend, for linear and head-to-tail peptides. We followed the released peptide-example schedule with 30 outer cycles, ProteinMPNN temperature 0.1, reference diffusion step 150 and `design_epoch_begin=20`. Cycles 1–20 evaluated one sequence per cycle, and cycles 21–30 evaluated two candidates. Target templates and target MSAs were supplied, with cyclic residue offsets enabled for head-to-tail designs. The native workflow has no hotspot conditioning, so hotspot requirements were applied during common evaluation. Both final cycle-30 candidates entered the native pool; intermediate-cycle candidates were excluded.

BoltzDesign1. We used the `release/v2` branch of the public BoltzDesign1 repository (Cho et al., 2025a) with its Boltz-2 backend for linear and head-to-tail peptides. The released protein-target recipe used 75 soft, 45 temperature and five hard optimization steps, SGD learning rate 0.2, Gumbel initialization and a helix-loss weight sampled from $[-0.3, 0]$. Target templates and unpaired target MSAs were supplied, with MSAs capped at 4,096 sequences. Hotspots entered native pocket conditioning, with `pocket_loss=False`. Backbones passing native ipTM > 0.5 underwent SolubleMPNN redesign, fixing interface residues within 4 Å and sampling 16 sequences at temperature 0.1; two sequences per backbone were retained for Boltz-2 reprediction. Native final outputs entered the pool with their pass flags. Gradient design, sequence redesign and native reprediction all counted toward the generation budget; common AF3/Protenix evaluation was separate.

Native pools are deduplicated by exact sequence and resolved topology within each method–case pair. External pools are uniformly sampled without replacement to $n = \min(N, 500)$ for common evaluation. HalluDesign and BoltzDesign1 remain below the cap. Both OmniPepBuilder arms contribute all native accepted candidates, as no case reaches 500. AF3 and Protenix evaluate the same selected cohort. Five diffusion samples constitute one candidate evaluation. Native filtering is part of each method's output definition; for OmniPepBuilder, the evaluation pool has already passed AF3-based generation acceptance.

Common AF3 evaluation uses version 3.0.4 with ten recycles; Protenix-v2 (Zhang et al., 2026b) uses ten cycles, 200 diffusion steps, bfloat16 precision and `reorder=false`. Both use the full prepared target assembly, observed target templates, no target MSA and five samples per candidate. The saved selected sample must satisfy at least one configured hotspot contact within 5 Å, each declared bond length within 0.3 Å of its nominal value, binder ipTM > 0.85 and peptide pLDDT > 0.80. Clash and target-reference RMSD thresholds are disabled. Confidence is tested on the saved selected prediction, without choosing a different sibling after that test. Accepted pre-Rosetta predictions are grouped by the target-frame GROMOS procedure, with backbone RMSD < 2.5 Å and length ratio > 0.8.

For native pool size N, evaluated cohort size n, accepted GROMOS group count D and retained generation exposure t, the main and Protenix figures report $(N/n)D/t$. This extrapolates observed accepted groups to the full native pool under a linear-growth assumption; observed-cohort throughput is D/t. Because GROMOS groups depend on the sampled set, new candidates can join existing groups or change their partition. At the fixed evaluation cap, faster generators can produce larger native pools and hence receive larger N/n multipliers. If additional candidates predominantly revisit existing groups, this linear expansion systematically overestimates their full-pool diversity and throughput. No cluster-count uncertainty interval is inferred. Generation exposure includes allocated startup and idle time and RF ProteinMPNN time; common reprediction, target-MSA preprocessing and discarded failed generation attempts are excluded. RF totals are 12.19–12.72 GPU-hours, and the eight shortened AnewOmni runs span 0.81–11.65 hours. Rates use actual retained exposure.

For ablations, removing revisit penalties sets $\alpha = 0$. The single-cell variant sets the pose-distance threshold to 10^6 Å and disables the length-ratio criterion, bringing sources into one spatial pool: Pareto competition broadens, and the common exposure penalty cancels from normalized weights. Removing replay restores independent fresh lineages. Every native candidate in these ablation cohorts is repredicted, without subsampling. At each native batch-completion time, accepted common-reprediction structures are grouped using only the available prefix. The 12-hour comparison therefore uses observed groups without linear expansion. All 16 linear/head-to-tail cases appear in the curves; the percentage summary excludes the two TNFα cases with zero full-policy yield. Ablations resume after approximately six hours, while controls run continuously; elapsed-time accounting excludes pauses. Each case is one seeded trajectory.

C.4 Additional benchmark results

Protenix reprediction retains passing designs from the full policy in 21 of 24 linear, head-to-tail and disulfide cases, with TNFα remaining the common zero-yield target (Fig. 14). Method rankings change relative to AF3. Of 3,566 full-policy candidates passing common AF3 evaluation, 1,536 also pass Protenix. Across all methods, no Asp–Lys candidate passes among 4,923 evaluated, and only two Glu–Lys candidates pass among 4,693. Saved bond-closure checks identify topology failures in all 163 full-policy MDM2 Asp–Lys candidates and all 500 BoltzGen MDM2 Glu–Lys candidates examined. The evaluation supplied the intended covalent connectivity. These outcomes emphasize predictor-specific confidence and topology handling in cross-model assessment.

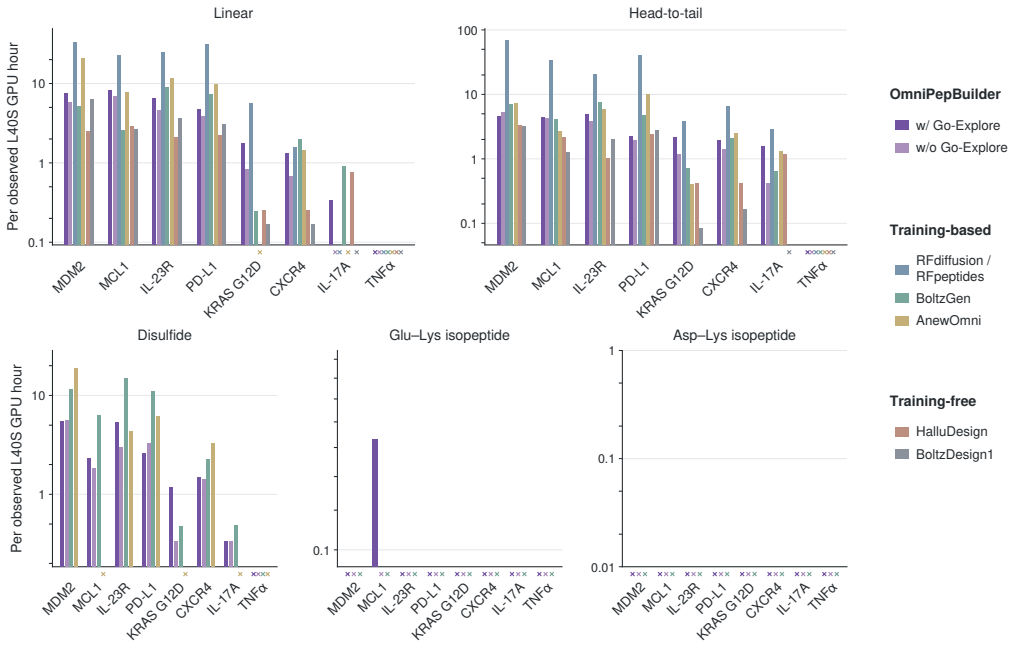


Figure 14: Global proposal benchmark evaluated by Protenix reprediction. Cohorts, acceptance gates, GROMOS grouping, and generation-cost accounting follow the AF3 benchmark, with Protenix replacing AF3 as evaluator. Values show $(N/n)D/T$ under the same hypothetical linear-growth assumption, not measured full-pool diversity. Logarithmic axes have independent ranges; crosses indicate no passing candidates. Absent methods were not benchmarked for that topology.

AF3 ipTM and Rosetta interface-energy distributions use the same 42,036 complete-metric candidates from 51,429 formal-cohort evaluations (Fig. 15; Fig. 16). They include confidence-rejected candidates with valid hard-gate and scoring outcomes. The omitted records comprise 8,736 without Rosetta scores, 619 failed pipeline/scoring cases and 38 anomalous scores below $-1{,}000$ REU. Retention varies by topology: 15,895/16,381 linear, 15,418/16,063 head-to-tail, 7,188/9,369 disulfide, 1,474/4,693 Glu–Lys and 2,061/4,923 Asp–Lys candidates. These distributions describe each method's native output pool after successful structural and scoring evaluation.

Rosetta uses full-complex FastRelax with the `beta_jan25` score function and a maximum of 200 relaxation iterations. The plotted interface score is raw separated-interface energy in Rosetta energy units, with no energy threshold in candidate acceptance. The paired figures expose confidence and interface-score variation beyond binary acceptance. Density normalization does not encode cohort size, and no native-pool expansion is applied.

Across the seven head-to-tail target cases with scored candidates from both OmniPepBuilder policies, median repredicted ipTM remains similar: 0.866–0.879 with replay and 0.863–0.885 without replay. The replay policy has a more favorable median Rosetta interface score in all seven cases. For example, median scores are -46.19 versus -42.72 REU for IL-17A and -56.57 versus -49.33 REU for CXCR4. Thus, the additional pose-group yield in these cases accompanies favorable interface scores while preserving high median reprediction confidence.

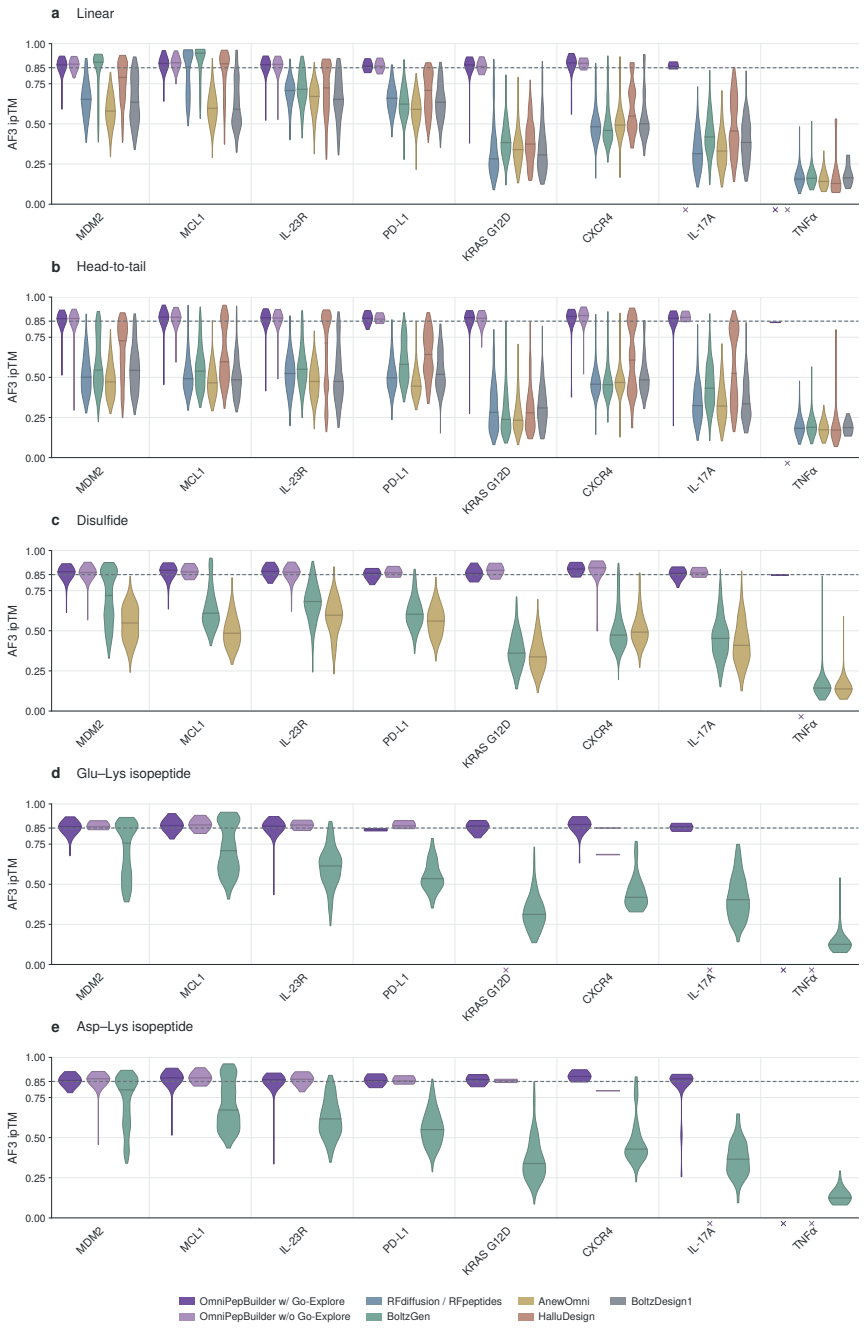


Figure 15: AF3 confidence distributions across five peptide topologies. Both score-distribution figures use the same 42,036 complete-metric records among 51,429 formal-cohort repredictions, including confidence-rejected candidates with otherwise valid scores. Exclusions comprise 8,736 unscored records, 619 failed pipeline/scoring cases, and 38 anomalous Rosetta scores below −1,000 REU. Dashed lines mark ipTM 0.85; short segments mark medians. Gaussian KDE bandwidth is 0.04 with reflected boundaries; one- or two-candidate groups use ticks. Crosses mark benchmarked groups without usable candidates. Violin width does not encode sample count.

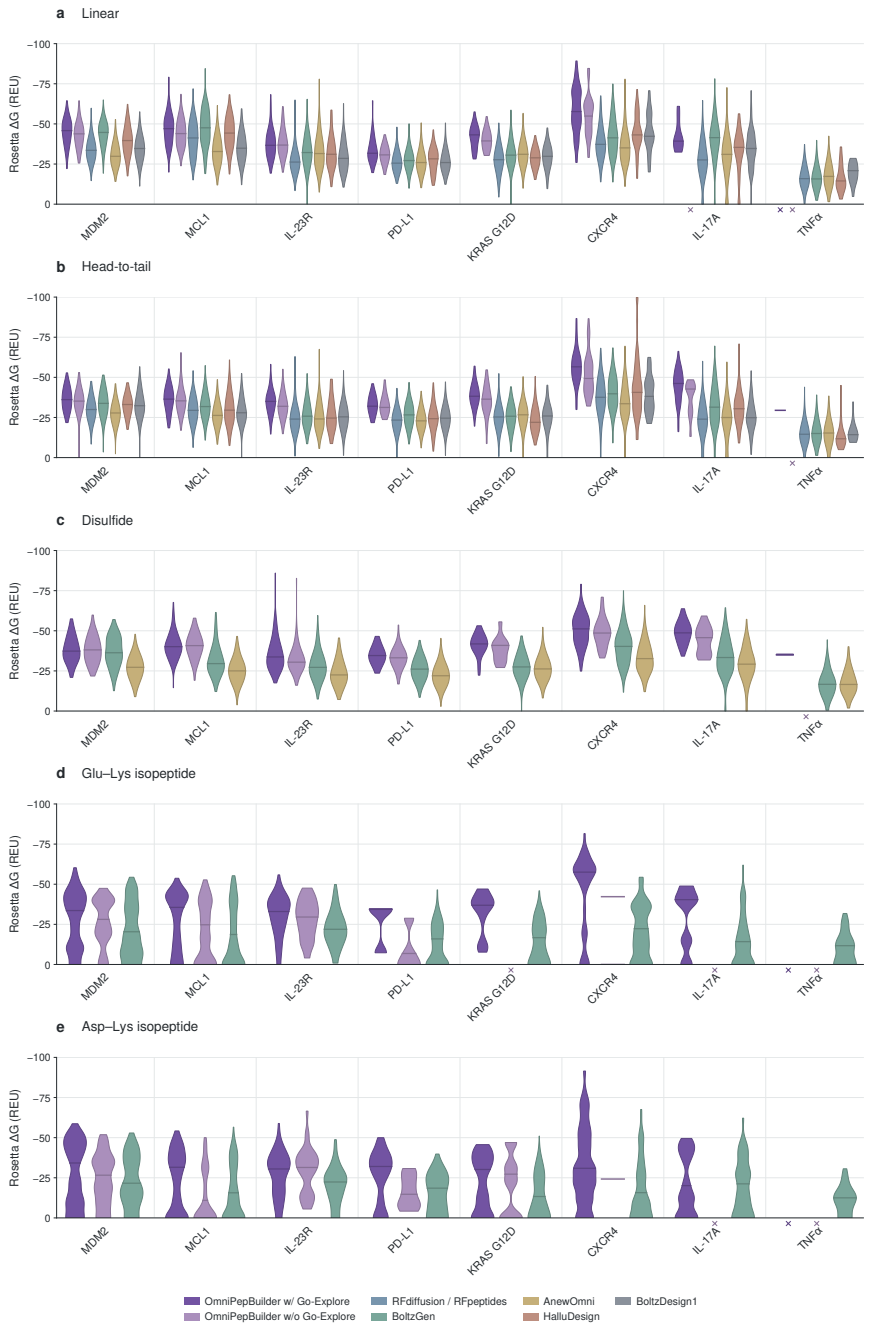


Figure 16: Rosetta interface-energy distributions for the same paired cohort as the AF3 confidence distributions. Values are raw dG_separated in REU; 347 positive values are capped at zero for density estimation. Gaussian KDE bandwidth is 4 REU, with boundary reflection and trimming to the observed range. Lower, more favorable scores appear higher on the plotted axis. Short segments mark medians, ticks show groups with fewer than three candidates, and crosses indicate groups without usable candidates. No sampling expansion or occurrence weighting is applied.

The ablation trajectories resolve differences hidden by the aggregate median (Fig. 17). Removing the revisit penalty increases head-to-tail yields for MDM2 (70 versus 62) and MCL1 (82 versus 66). Retaining it increases yields for both KRAS G12D topologies, head-to-tail CXCR4 (35 versus 26) and head-to-tail IL-17A (21 versus 19). TNFα linear remains zero in all arms; TNFα head-to-tail yields zero groups with the full and no-replay policies, one without the revisit penalty and three with a single cell. The curves show how the allocation tradeoff varies across targets and topologies over the common observation window.

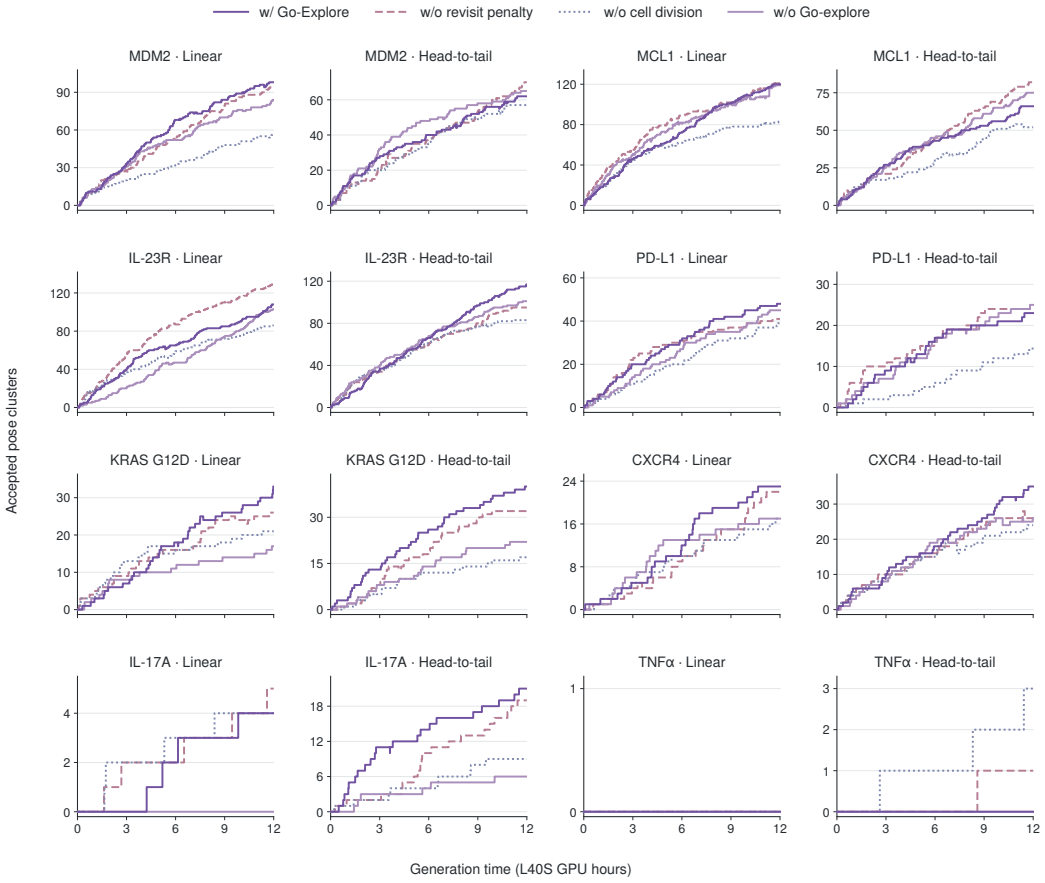


Figure 17: Accepted pose groups after AF3 reprediction over 12 active L40S generation GPU-hours. Curves retain all 16 target–topology cases, including the two zero-reference TNF-alpha cases omitted from percentage-change summaries. GROMOS is recalculated at each accepted prefix, so counts can decrease. Pauses and reprediction runtime are excluded. These are single seeded trajectories; ablations resumed while controls ran continuously. Coincident curves overlap.

C.5 ADDITIONAL VISUALIZATIONS

The TNFα WP9 examples follow extended disulfide and head-to-tail searches, separate from the 12-hour benchmark. Each uses four GPUs and terminates by cancellation with 13 accepted unique sequences in three retrospective native-pose groups. The plotted time includes execution startup and generation overhead, excluding queue time and inter-job gaps. Native acceptance requires ipTM ≥ 0.85, peptide pLDDT ≥ 0.80, three specified hotspot contacts and valid topology. GROMOS is recalculated at every accepted prefix.

These trajectories used the earlier cell-based allocator, with conditional cell weights $\exp(-F_c/0.1)/\sqrt{1+N_{\mathrm{return},c}}$; the current exact-source formulation is given in Appendix C.1. The confidence-plane paths track the destination cells receiving the first accepted members of groups G1–G3. Each coordinate is a cellwise maximum, so the plotted ipTM and pLDDT can come from different states. Filled markers identify first native acceptance, and hollow markers identify cell birth. Color denotes the conditional return probability over the active archive, with probabilities above 8% saturating the common color scale. Changes reflect confidence, revisit counts and normalization over the evolving pool.

For Fig. 18, we select one strictly passing AF3-repredicted member per native group by highest ipTM, then peptide pLDDT and candidate ID. The selected models satisfy the saved hotspot/topology gates and ipTM ≥ 0.85, peptide pLDDT ≥ 0.80. Their group identities refer to native discovery rather than a new partition of repredicted poses. All six pre-Rosetta structures are aligned through 456 target Cα atoms and shown using the same view and scale. The gold hotspot residues L75, I97 and N137 use prepared-reference numbering; they correspond to positions 70, 92 and 132 in the predicted chains. The illustrated structures provide sequence-level reprediction support and visualize alternative target-relative poses.

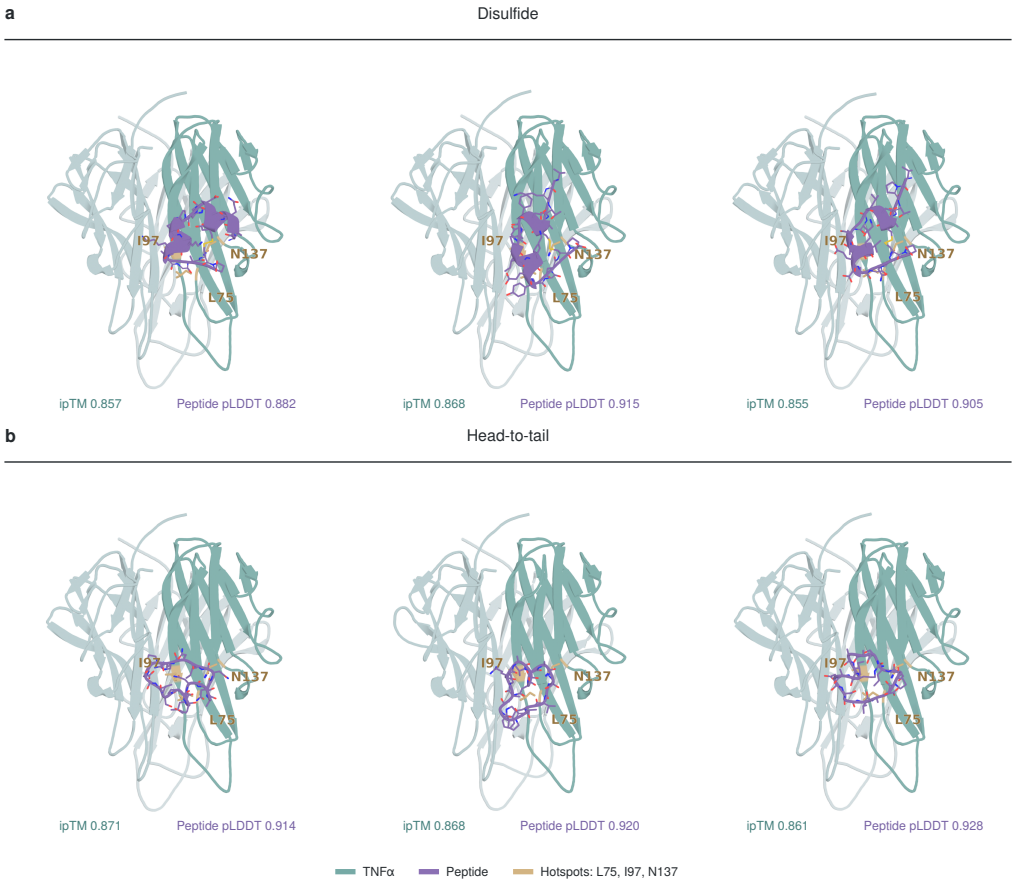


Figure 18: AF3 repredictions of selected representatives from TNF-alpha native search. Rows show disulfide and head-to-tail designs from the three native pose groups. These selected representatives illustrate repredicted structures and confidence; native group membership does not imply that every member passes common repre-diction.

D LOCAL REFINEMENT

D.1 METHOD DETAILS

Search space and mutation proposals. Each search starts from a fixed target–peptide complex with sequence s_0 and reference pose X_0. Peptide length, covalent topology and immutable residue positions are preserved. We permit at most five substitutions relative to s_0, using the residue numbering of the starting complex. All non-fixed positions are mutable in the reported experiments. Proposals change one position at a time and exclude cysteine from the substitution alphabet. A supplied interface-distance cutoff can instead restrict the mutable positions to a mask computed once from the initial complex.

Let $\mathcal{N}_5(s)$ be the legal one-substitution neighbors of s within $\mathcal{B}_5(s_0)$. Both controllers use

$$q(s' \mid s) = 0.20\,\frac{1}{|\mathcal{N}_5(s)|} + 0.80\,\frac{p_{\mathrm{MPNN}}(s'_i \mid X_s, s_{-i})}{\sum_{z\in\mathcal{N}_5(s)} p_{\mathrm{MPNN}}(z_{i(z)} \mid X_s, s_{-i(z)})}, \quad s' \in \mathcal{N}_5(s), \tag{5}$$

where i is the changed position and X_s is the relaxed structure of the current sequence. ProteinMPNN uses the `v_48_020` model in conditional-probability mode (Dauparas et al., 2022); its probabilities are renormalized over legal substitutions. At the radius boundary, outward moves are removed while reversions and substitutions at already changed positions remain available. The uniform component gives every legal substitution positive probability.

Structural eligibility and energy evaluation. Each starting sequence and newly proposed sequence receives an independent structure prediction with five diffusion samples. A sample must jointly satisfy binder ipTM > 0.85, peptide pLDDT > 0.80, and peptide Cα RMSD < 2.5 Å relative to X_0. RMSD uses target-only alignment without a second peptide fit. The alignment support is the fixed target binding patch, defined by target Cα atoms within 10 Å of a hotspot Cα; without hotspots, the whole target supplies the alignment. Equivalent target-chain assignments are handled consistently. Cyclized candidates also satisfy the configured bond-length tolerance, ± 0.30 Å around the expected linkage distance. The highest native-ranking sample among those passing all gates is selected.

Rosetta relaxes this selected prediction with full-complex FastRelax, the `beta_jan25` score function and a maximum of 200 minimizer iterations. Peptide and target backbone and side-chain torsions may move; rigid-body jumps remain fixed. Bond-bearing topologies use the corresponding constrained score-function variant. InterfaceAnalyzer separates the peptide from the complete target group, with `pack_input=False` and `pack_separated=True`. The search energy is the resulting `dG_separated` in REU, and the root-relative objective is $F(s) = E(s) - E(s_0)$. The local-refinement protocol additionally requires target-aligned peptide Cα drift below 1.5 Å between the selected prediction and its relaxed output. The energy is evaluated on that relaxed structure; predictor confidence remains attached to the selected prediction.

Evolutionary search. The population initially contains the valid starting sequence and grows to at most four parents. Each iteration draws eight new children across the population through round-robin parent traversal. For each parent, directed mutation edges are drawn without replacement using the restricted proposal distribution. After all children in the batch finish evaluation, the current parents and valid children form the survivor pool. The three lowest-energy sequences are retained as elites; the remaining population slot is sampled uniformly from non-elites. Selection uses $F(s)$, with sequence order resolving exact ties. There is no crossover. A final partial batch is evaluated before termination.

Algorithm 2 Basin-constrained evolutionary local refinement

Require: Starting complex (s_0, X_0); radius 5; budget B; population 4; offspring batch 8; elites 3.

1: Predict and gate s_0 against X_0; relax and score the passing representative.
2: If root evaluation fails eligibility, stop this root; otherwise set $\mathcal{P} = \{s_0\}$.
3: Initialize sequence cache $\mathcal{C}$, evaluated archive $\mathcal{V}$, and counter $b = 0$.
4: **while** $b < B$ and an unseen child is reachable from $\mathcal{P}$ **do**
5: Traverse parents round-robin; draw untried legal mutation edges using Eq. 5.
6: Reserve up to $\min(8, B - b)$ uncached children; increment b for each reservation.
7: **for each** reserved child s **do**
8: Predict five samples; select the highest-ranked jointly passing sample.
9: If eligible, relax, score, and check relaxation drift; record the result in $\mathcal{C}$.
10: For valid s, store $F(s)$ and conditional ProteinMPNN probabilities in $\mathcal{V}$.
11: **end for**; wait for the complete offspring batch.
12: Pool current parents and newly valid children; retain the best three as elites.
13: Fill remaining population slots uniformly from non-elites, up to four parents.
14: **end while**; retain every valid archived s with $F(s) < 0$.
15: After all roots finish, merge eligible cyclic rotations; return the minimum-E representative of each identity.

Metropolis/annealing. Eight walkers share the evaluated archive and sequence cache. For a fully evaluated, valid proposal s' from s, the acceptance probability includes the asymmetric proposal correction:

$$a(s, s') = \min\left\{1, \exp\left[-\frac{F(s') - F(s)}{T}\right] \frac{q(s \mid s')}{q(s' \mid s)}\right\}. \tag{6}$$

Each walker cools geometrically from 1.0 to 0.25 REU over 24 completed attempts. Invalid proposals are rejected and count as attempts. After 24 attempts without a new personal best, the walker restarts from the lower-energy member of a two-draw tournament over favorable archived candidates, falling back to the initial sequence when none exist, and resets its temperature. Valid candidates remain in the archive even when a walker rejects their proposed transition.

Caching, budgets and cyclic identity. Exact sequence strings share one evaluation cache per root. A first reservation consumes budget even if the candidate subsequently fails a structural gate; completed or in-flight cache hits do not launch duplicate evaluations. The default production limit is 500 non-root reservations per starting complex. Final eligibility uses all valid evaluated sequences with $F(s) < 0$, independently of membership in the final population or walker states.

For head-to-tail peptides, the search retains the original residue registration for Hamming distance, structure coordinates and ProteinMPNN probabilities. After all roots finish, accepted candidates are grouped by cyclic sequence rotations together with their covalent annotations. Reverse sequence order is distinct. Among equivalent molecular identities, the representative with the lowest raw Rosetta dG is retained; a stable candidate identifier resolves ties. This prevents redundant final candidates while preserving the position-specific search.

D.2 Benchmark results

Calibration settings. We compared six accepted linear starting complexes targeting MDM2, PD-L1, IL23R, KRAS G12D, IL17A and CXCR4. Three campaign seeds (2026, 12026 and 22026) were used. CXCR4 seed 12026 failed the initial prediction gates in both arms and was excluded, leaving 17 paired target–seed blocks. Predictions used Protenix v1 (`protenix_base_default_v1.0.0`) (Zhang et al., 2026c), target-only templates, no MSA, ten recycling cycles, 200 diffusion steps and five samples. The controller-calibration runs applied the confidence and root-relative pose gates; the two subsequent case-study runs additionally applied the 1.5 Å relaxation-drift gate.

The coarse calibration comprised three rounds. First, Metropolis initial temperatures of 1, 2 and 4 were compared, while evolutionary population/elite sizes were (4,2), (8,4) and (16,8). Second, archive-restart probabilities of 0, 0.5 and 1 were compared for Metropolis, and elite sizes of 1, 2

and 3 at population size four for evolutionary search. Third, ProteinMPNN proposal fractions of 0, 0.5 and 0.8 were compared for each retained controller. The first round used 192 reservations per run. Later Metropolis rounds used 384 reservations to include post-restart exploration; evolutionary rounds retained 192. These comparisons selected the settings described above. The 500-reservation production cap is a practical budget choice beyond these calibration horizons.

Matched controller comparison. Fig. 19 compares the selected controllers over the first 192 reserved non-root sequences in each run. Valid-candidate score distributions use only completed candidates passing the structural gates; valid and root-improved fractions retain all 192 reservations in their denominator. Run-level comparisons are paired by target and seed. Evolutionary search produces a lower run-median dG in 15/17 pairs, with a median paired improvement of 2.42 REU. The median paired difference favors evolutionary search for every target (Table 4). Averaging the per-target mean fractions, valid yield is 62.9% versus 56.0%, and yield with dG below the run's own root is 48.8% versus 36.4%. This equal-evaluation comparison motivates evolutionary search as the default for our local-refinement stage.

Table 4: Comparison at 192 reserved evaluations per run. Differences are evolutionary minus Metropolis run-median dG, then summarized by the median across paired seeds for each target; negative values favor evolutionary search.

Target	Paired runs	Median dG difference (REU)
MDM2	3	-8.89
PD-L1	3	-1.13
IL23R	3	-3.92
KRAS G12D	3	-0.54
IL17A	3	-2.56
CXCR4	2	-4.25

Selected interface refinements. The two examples in Fig. 6 start from AF3-generated accepted complexes and use the calibrated evolutionary controller with a 500-reservation budget. For KRAS G12D, the sequence changes from `PTSEERIEAIRNL` to `MDSLERQEAIRFL` through P1M, T2D, E4L, I7Q and N12F. The selected candidate has target-aligned prediction RMSD 1.80 Å and dG improvement 12.89 REU. Phe12 adds an aromatic side chain between target His95 and Tyr96, increasing the local packing contacts highlighted in the figure.

For head-to-tail CXCR4, `LPNGRGIG` becomes `YPPGVGIG` through L1Y, N3P and R5V. Its target-aligned prediction RMSD is 2.45 Å and dG improvement is 19.79 REU. In the retained relaxed structure, Tyr1 expands the side-chain contact neighborhood to include Tyr166, Phe175, Gln176, His179 and Tyr231 (model residue numbering), compared with Gln176 alone for the initial Leu1 at the same 4 Å heavy-atom cutoff. The pre/post complexes retain the head-to-tail amide closure. These structural comparisons describe the jointly optimized candidates, whose high-confidence predicted poses permit further local packing improvements.

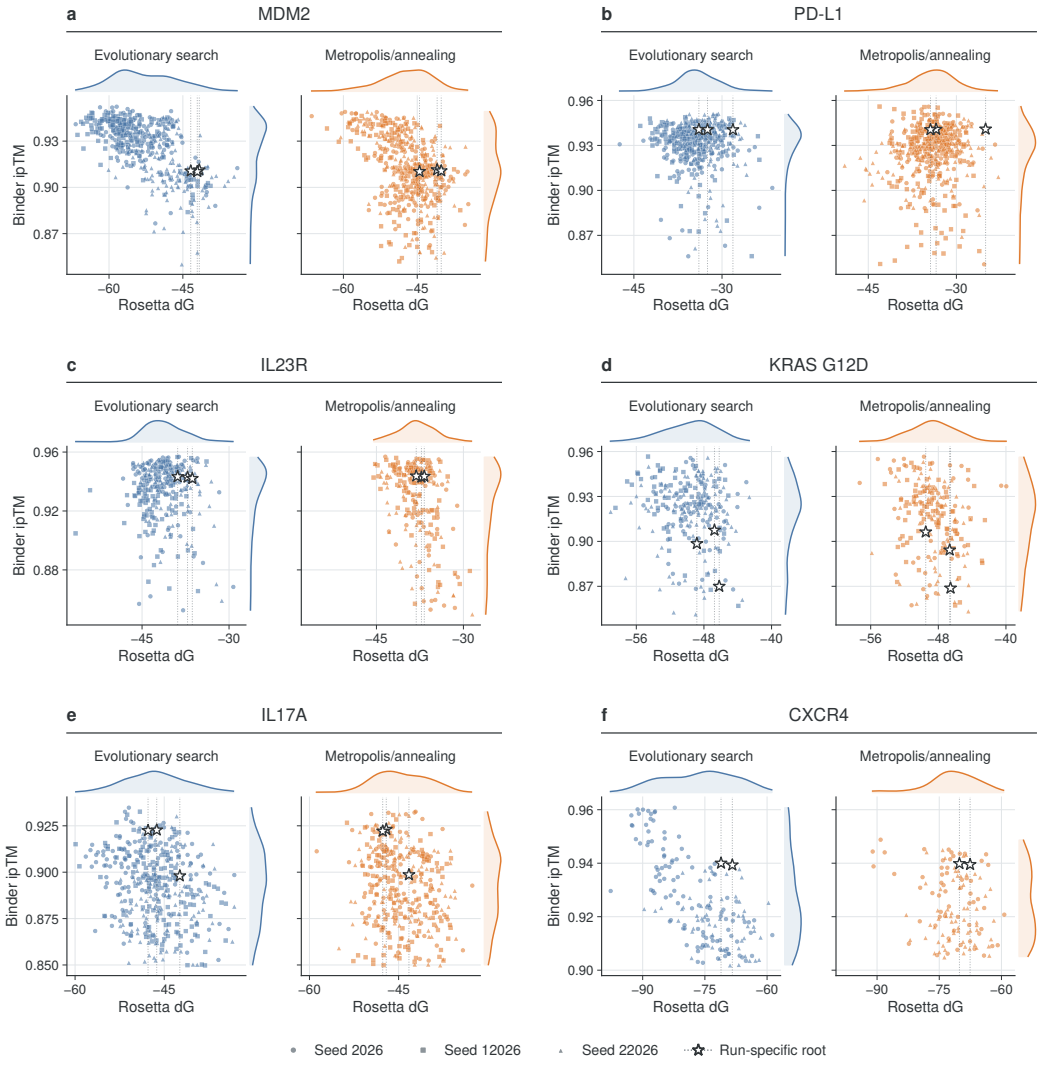


Figure 19: Evolutionary search and Metropolis/annealing across six targets and 17 matched target–seed blocks (34 runs). Both controllers are restricted to their first 192 reserved non-root sequences; only valid completed candidates are plotted. White stars and dotted lines mark each run's root, and marker shapes identify seeds. Axes are matched within target pairs. Rosetta dG is the separated interface-energy score in REU; binder ipTM is from Protenix. This compares a common reservation budget, not production-horizon throughput.

E Designing Multi-target Binders and Macrocyclic Glues

E.1 Method details

Physical task representation. A multi-target candidate consists of one peptide sequence, one resolved covalent topology, and a separate target–peptide complex for each of K required targets. The peptide can adopt different bound conformations across these complexes. Peptide length, fixed residue identities and linkage endpoints are resolved once and shared across all target evaluations. A glue candidate instead consists of one physical ternary complex containing the peptide and both protein partners. All interface measurements for a glue candidate come from the same selected diffusion sample. Thus, independent multi-target binding and simultaneous ternary bridging use distinct structural representations within the same search framework.

Coupled sequence proposals. Multi-target ProteinMPNN proposals use the `v_48_020` checkpoint (Dauparas et al., 2022) with one structural graph per target complex. Every graph has the same peptide decoding order and partial peptide sequence. At position i, target-specific logits $z_{t,i,a}$ for amino acid a are averaged before sampling:

$$\bar{z}_{i,a} = \frac{1}{K}\sum_{t=1}^{K} z_{t,i,a}, \qquad p(a \mid i) = \frac{\exp(\bar{z}_{i,a}/\tau)}{\sum_{b\in\mathcal{A}_i}\exp(\bar{z}_{i,b}/\tau)}, \quad a \in \mathcal{A}_i, \tag{7}$$

where $\tau = 0.1$ and $\mathcal{A}_i$ is the allowed residue alphabet. One sampled identity is fed back to every target's decoder before the next position is generated. Target sequences and topology-required peptide residues remain fixed. This gives a shared sequence proposal conditioned on every selected pose, while keeping the independent complexes geometrically separate. For glue candidates, ordinary ProteinMPNN sampling conditions on the entire ternary structure, with both target sequences fixed. New unconstrained cysteines are excluded in both modes; the resolved disulfide endpoints remain cysteine.

Joint pose neighborhoods, Pareto retention and replay. A state is represented by its peptide pose in each required target frame. For multi-target design, these frames come from separate predictions; for glue design, they come from the same ternary prediction. Each target is aligned to its own fixed reference patch, using the global-proposal correspondence and symmetry rules (Appendix C). Two states are direct neighbors only when a common, direction-preserving peptide registration gives backbone RMSD below 2.5 Å in every target frame and their peptide-length ratio exceeds 0.8. Cyclic registrations are evaluated jointly across frames. No additional peptide fit or complex TM-score threshold is used. This prevents a match at one interface from hiding a different pose at another.

Local Pareto retention compares directly matching states with the same restart constraints. A multi-target state contributes $2K$ confidence coordinates, $(I_1, P_1, \ldots, I_K, P_K)$, where I_t and P_t are peptide–target ipTM and peptide pLDDT in complex t. A two-partner glue contributes (I_1, I_2, P) because it has two peptide–partner interfaces but one physical peptide. Retained states are partitioned by GROMOS using the joint pose-match graph. With confidence values normalized to 0–1, allocation uses

$$\begin{aligned} E_s^{\text{multi}} &= \max\Big\{0, \max_t(0.85 - I_{s,t}), \max_t(0.80 - P_{s,t})\Big\}, \\ E_s^{\text{glue}} &= \max\{0, 0.85 - I_{s,1}, 0.85 - I_{s,2}, 0.80 - P_s\}. \end{aligned} \tag{8}$$

These energies enter the state-selection distribution in Eq.˜2, retaining the cell-size correction, direct-neighbor revisit exposure, temperature 0.10 and revisit exponent 0.5. One return allocation restores the complete state and incurs one revisit charge. No target–target ipTM is added to the Pareto vector or allocation objective.

Prediction and acceptance. The application campaigns use AF3 3.0.4 with target templates, no target MSA, ten recycles and five diffusion samples per physical prediction. Multi-target branches select their representatives independently, using the highest native-ranking sample among hard-gate survivors. Glue selection prioritizes the minimum peptide–partner ipTM among hard-gate survivors. A candidate can continue only when every required target passes its hard gates, and is accepted only when every peptide–target ipTM is at least 0.85 and every peptide pLDDT is at least 0.80.

The contact gate requires at least three contacted hotspot residues per target, using a 5 Å heavy-atom cutoff. Where no positive hotspots are specified, it instead requires three contacted residues anywhere on that target. Any contact within 5 Å of a designated coldspot rejects the sample. Specified cyclization bonds must lie within ± 0.30 Å of the expected bond length. Clash counts are recorded, but clash rejection and target-reference RMSD rejection were disabled in these campaigns. Glue target–target contacts are measured without imposing a minimum. Each multi-target candidate is accepted, archived or continued only after all of its target evaluations have completed.

Design tasks and target context. Table 5 lists the seven campaigns and their retained outcomes. MDM2/MDMX uses the p53-binding domains with 8–18-residue head-to-tail peptides. GLP1R/GCGR and GLP1R/GCGR/GIPR use 24–36-residue linear peptides, with active-state receptor references from PDB 6X18, 6LMK and 7DTY, respectively. SSTR2/SSTR5 uses 8–18-residue peptides with one disulfide bond; its cysteine endpoints are sampled at initialization and then preserved. Positive hotspot sets direct these searches to the corresponding peptide-binding pockets.

The glue tasks use VHL residues 54–213, BRD4 BD1 residues 42–168, KRAS G12D residues 1–169, or the KLHL20 Kelch domain, residues 303–605. All three use 8–18-residue head-to-tail peptides. Neither partner has positive hotspot restrictions, allowing the search to discover both peptide placement and relative protein orientation. VHL coldspots exclude 28 residues in the reviewed Elongin/Cullin interface footprint; KRAS coldspots exclude residues 15, 16, 18, 116 and 146 near the GDP-occluded cavity. AF3 inputs contain the selected protein components and peptide; nucleotide, membrane and the remaining E3 assembly components are not included. The reported experiments comprise global proposal without subsequent local refinement or Rosetta selection.

Search settings and campaign outcomes. Each case uses seed 2026, eight active lineages, an eight-cycle rollout cap and return probability 0.70. Random starts exclude cysteine except where required by the topology. The origin and accepted-candidate limits were set to 10^6 so that elapsed time determined the practical campaign budget. Every campaign used four NVIDIA L40S GPUs and completed the reported search interval in less than 24 hours. Logical candidate counts refer to shared molecules with complete required-target evidence; a multi-target candidate entails K separate prediction ensembles.

Table 5: Application campaigns on four L40S GPUs per case. Accepted counts follow the complete computational acceptance policy. Pose groups are accepted-only GROMOS clusters under the joint matching rule.

Case	Topology	Length	Time (h)	Accepted	Groups
MDM2/MDMX	Head-to-tail	8–18	5.34	134	42
GLP1R/GCGR	Linear	24–36	23.99	16	8
SSTR2/SSTR5	Disulfide	8–18	19.33	7	5
GLP1R/GCGR/GIPR	Linear	24–36	16.75	0	0
VHL–BRD4	Head-to-tail	8–18	11.31	73	19
VHL–KRAS G12D	Head-to-tail	8–18	23.97	2	1
KLHL20–BRD4	Head-to-tail	8–18	19.15	23	6

Representative selection. Accepted candidates are grouped using the saved joint pose-match graph. Within each group, the representative maximizes the minimum peptide–target ipTM, followed by minimum peptide pLDDT and a stable candidate identifier. The highest-ranked representative supplies each main-text example, except GLP1R/GCGR: the third-ranked candidate was selected after the first two showed severe short contacts. The selected sequences are `VARFFGLPGIP` (MDM2/MDMX), `GLGEAEFFNELDKLLDRKAALEFIQRLIELSEQ` (GLP1R/GCGR), `PVGGCDPYGNCY` (SSTR2/SSTR5), `GGVGTGLEG` (VHL–BRD4), `PNSSRPVPADKVHLY` (VHL–KRAS G12D) and `GPLGGKYN` (KLHL20–BRD4). The SSTR peptide has a Cys5–Cys11 disulfide; the other cyclic examples are head-to-tail amides. All displayed structures are retained AF3 predictions. The main MDM2/MDMX, GLP1R/GCGR and SSTR2/SSTR5 representatives contain 11, 33 and 12 residues, with paired ipTM values of 0.917/0.918, 0.864/0.887 and 0.904/0.919, respectively. The VHL–BRD4, VHL–KRAS G12D and KLHL20–BRD4 representatives contain 9, 15 and 8 residues, with minimum peptide–partner ipTM values of 0.917, 0.870 and 0.904.

E.2 Additional visualizations

Multi-target recognition. Fig. 20 presents three additional candidates for each dual-target case and two exploratory triple-target candidates. Additional examples are selected for pose variation after confidence ranking and a 2 Å inter-chain heavy-atom distance screen, with the two stated exceptions below. For visualization selection, structural separation is measured by a symmetric nearest-neighbor peptide Cα distance after target alignment, averaged across target branches. This accommodates differing peptide lengths; campaign pose-group counts instead use the registered backbone RMSD criterion described above.

MDM2/MDMX and SSTR2/SSTR5 admit visibly different peptide arrangements, whereas the GLP1R/GCGR examples predominantly vary a helical binding arrangement. The amber pocket annotations connect these predictions to known recognition differences. In the main MDM2/MDMX example, Phe5 contacts Leu54/Met53 at 3.92/3.71 Å and Phe4 contacts His96/Pro95 at 3.55/3.53 Å, accommodating distinct side-chain environments in the p53-binding pockets (Pazgier et al., 2009). For GLP1R/GCGR, the modeled Phe8 contacts Lys197/Ile194 and the Gly1 backbone contacts Val237/Ile235, positions implicated in receptor-dependent recognition of known dual agonists (Li et al., 2023). In SSTR2/SSTR5, the constrained loop contacts Thr194/Asn187 and Phe294/Tyr286; the latter pair has experimental support for receptor-dependent octreotide responses (Li et al., 2024b). These annotations identify predicted contacts in pharmacologically relevant pockets, without assigning the designed peptides an experimentally established signaling function.

The two GLP1R/GCGR/GIPR examples approach the all-target confidence thresholds but remain unaccepted. For the first, ipTM values across GLP1R/GCGR/GIPR are 0.866/0.840/0.847, with peptide pLDDT 0.835/0.794/0.822. For the second, the corresponding values are 0.846/0.847/0.837 and 0.819/0.798/0.814. Their maximum confidence deficits are 0.0104 and 0.0127. They illustrate the tighter intersection imposed by a third receptor rather than accepted triple-target designs. Two selected structures retain short contacts: GCGR Gln27 N–peptide Asn14 ND2 at 1.925 Å in the right GLP1R/GCGR example, and GLP1R Tyr152 OH–peptide Glu3 OE1 at 1.967 Å in the left triple-target example.

Ternary-complex geometries. Fig. 21 shows three additional head-to-tail macrocycles for each E3–target pairing, with both a surface overview and an interface closeup. The VHL–BRD4 and KLHL20–BRD4 examples vary peptide placement and the orientation of BRD4 relative to the E3 component, illustrating alternative assemblies discovered without a fixed inter-protein arrangement. All nine displayed additional ternary models pass the 2 Å inter-chain distance screen, including the target–target interface.

For VHL–KRAS G12D, the first additional example is the second accepted candidate. The other two are retained alternatives with minimum peptide–partner ipTM of 0.833 and 0.835, below the 0.85 acceptance threshold. Both accepted VHL–KRAS candidates belong to the same accepted pose group; the lower-confidence examples expand the structural comparison without being counted as accepted outcomes. These models assess simultaneous interface compatibility, rather than ternary cooperativity or degradation activity.

The peptide is violet in every panel; E3 components are sand and recruited partners teal. Interface closeups show target residues covering at least 80% of peptide–target heavy-atom pairs within 4 Å, retaining ties. These distances determine display coverage rather than energetic contributions or hydrogen-bond assignments. Camera alignment and surface representations preserve the predicted atomic coordinates.

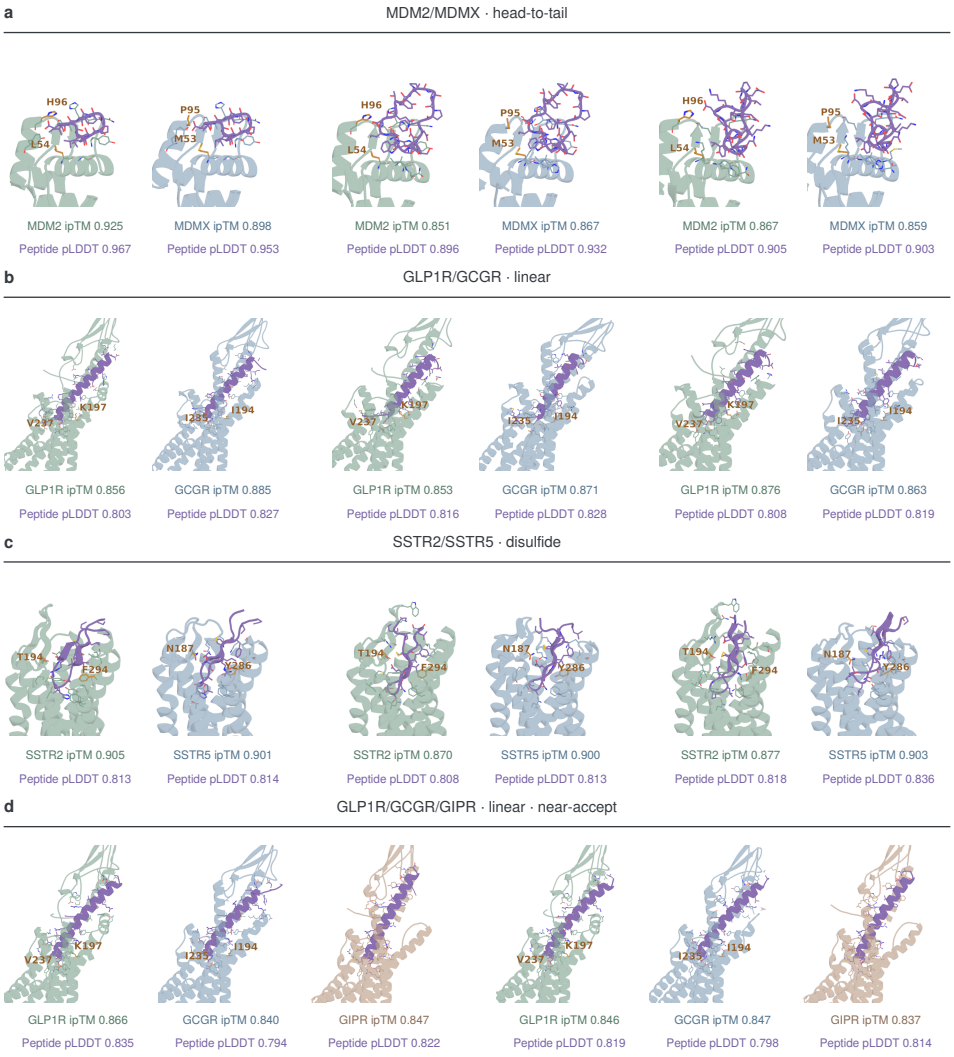


Figure 20: Additional multi-target peptide candidates. Three candidates each are shown for MDM2/MDMX, GLP1R/GCGR and SSTR2/SSTR5, followed by two GLP1R/GCGR/GIPR candidates. Each peptide is shown in separate target complexes; ipTM and peptide pLDDT are normalized to 0–1. Amber residues mark literature-supported pocket differences. Both triple-target examples in d are below acceptance criteria. The right example in b and the left example in d retain the documented 1.925 Å and 1.967 Å short contacts, respectively.

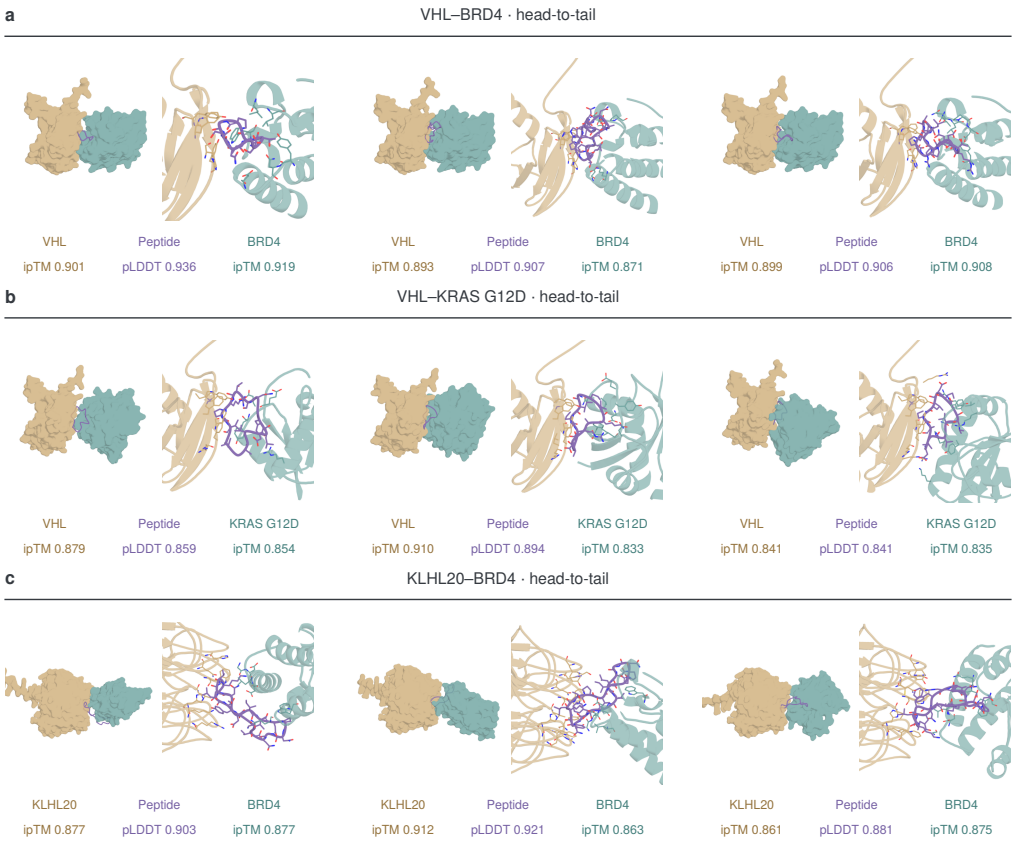


Figure 21: Additional macrocyclic glue candidates. Three designs each are shown for VHL–BRD4, VHL–KRAS G12D and KLHL20–BRD4, with a surface overview and interface closeup per candidate. Peptides are violet, E3 components sand and their partners teal. Labels give peptide–target ipTM and peptide pLDDT normalized to 0–1. The middle and right examples in b are below the confidence acceptance criteria, with minimum ipTM 0.833 and 0.835. Structures and scores are predictions.