
IDPepDesign: Insertion-Deletion is all peptide design needs

Anonymous Author(s)

Affiliation

Address

email

Abstract

In protein binder design, many promising candidates are too large for efficient synthesis, expression, or therapeutic delivery. Converting these binders into short peptides while preserving binding affinity and specificity typically requires extensive experimental optimization. At the same time, designing peptide binders directly from a target protein sequence remains substantially more challenging than unconditional peptide generation, as the model must simultaneously capture the distribution of short peptides and learn target-specific interactions. Here, we present IDPepDesign, a target-conditioned peptide design framework built on SCISOR, an insertion–deletion diffusion model for protein sequence generation. IDPepDesign introduces two complementary conditioning mechanisms: a pooled target representation injected throughout the denoising process, and residue-level cross-attention that enables every peptide position to attend to encoded target residues. This architecture transforms SCISOR from an unconditional peptide generator into a target-aware binder design model while retaining its ability to generate realistic short peptide sequences. Our model substantially improved peptide generation quality (perplexity $17.1 \rightarrow 11.3$; Fréchet protein distance $1.84 \rightarrow 0.75$) and enabled target-conditioned peptide design. Across 11 established protein binder targets, IDPepDesign produced peptide binders with competitive predicted interface quality compared with both sequence- and structure-based design baselines.

1 Introduction

Peptide binders offer an attractive alternative to large protein binders because they are easier to synthesize, manufacture, and deliver while retaining the ability to recognize extended protein surfaces that are often inaccessible to small molecules. However, designing peptide binders that preserve the affinity and specificity of larger binders remains challenging, particularly for targets with limited structural information.

Structure-based design frameworks such as RFdiffusion with ProteinMPNN [Watson et al., 2023, Dauparas et al., 2022] and BindCraft [Pacesa et al., 2025] generate interfaces directly in three-dimensional space. While highly successful, these approaches inherit the Protein Data Bank’s bias toward well-folded, crystallizable complexes [Alamdari et al., 2023], leaving many therapeutically relevant targets—including intrinsically disordered proteins, transcription factors, and fusion oncoproteins—underrepresented. Sequence-based approaches avoid this limitation. PepMLM reconstructs peptide binders conditioned on target sequences using a span-masked ESM-2 model [Chen et al., 2025], while Forge employs latent flow matching over large protein interaction graphs to generate target-specific binders [Ko and Wang, 2025]. However, these methods are designed for de novo peptide generation rather than compressing an existing binder into a shorter peptide while preserving the key interface residues responsible for target recognition. SCISOR [Baron et al., 2025] provides a promising foundation for this task. Instead of masking residues as in conventional language models,

38 SCISOR learns an insertion–deletion diffusion process that inserts random amino acids during the
39 forward process and removes them during denoising. The same model can therefore both generate
40 sequences from noise and progressively shorten existing proteins, providing an inductive bias for
41 sequence compression that conventional autoregressive or masked language models do not possess.
42 However, SCISOR is an unconditional generative model trained on general protein sequences and
43 lacks any mechanism for conditioning sequence generation on a target protein.

44 Here, we introduce IDPepDesign, a target-conditioned extension of SCISOR for peptide binder design.
45 Our method augments the pretrained diffusion model with a frozen ESM-2 target encoder [Lin et al.,
46 2023], a lightweight trainable target transformer, global target conditioning, and residue-level peptide-
47 to-target cross-attention integrated throughout the denoising network. These conditioning pathways
48 are zero-initialized, allowing the pretrained SCISOR model to be preserved at initialization while
49 enabling target-specific interactions to be learned during fine-tuning. It combines SCISOR’s ability to
50 compress protein sequences with explicit target conditioning, enabling sequence-only peptide binder
51 design without sacrificing the underlying generative capabilities of the pretrained model.

52 1.1 Contributions

53 1) We propose IDPepDesign, a target-conditioned extension of SCISOR that enables sequence-
54 only peptide binder design. 2) We show that target-conditioned fine-tuning improves short-peptide
55 generation while preserving the pretrained generative capabilities of SCISOR. 3) We demonstrate
56 competitive peptide binder design across 11 benchmark targets, outperforming unconditional SCISOR
57 and random shortening baselines.

58 2 Method

59 2.1 Target Conditioning

60 Unlike conventional conditional generation with discrete class labels, peptide binder design requires
61 conditioning on previously unseen target proteins. Inspired by text-to-image diffusion models, we
62 condition SCISOR on continuous target representations extracted from a pretrained protein language
63 model [Lin et al., 2023]. During unconditional generation, these representations are replaced by
64 zeros. A target sequence is encoded with a frozen ESM-2 35M model, producing both residue-level
65 embeddings and a pooled sequence representation. These features are incorporated into SCISOR
66 through two complementary conditioning pathways (Figure 2): a global pathway that conditions
67 sequence generation on the target identity, and a residue-level pathway that enables peptide residues
68 to attend to specific target residues. The full architecture, is shown in Appendix A.

69 2.1.1 Pooled Pathway (global)

70 The pooled target embedding is projected into SCISOR’s global conditioning space and added to
71 the diffusion-time embedding. The final projection layer is zero-initialized so that the pretrained
72 SCISOR model is exactly recovered before fine-tuning.

73 2.1.2 Per-residue pathway

74 A lightweight trainable encoder refines the residue embeddings into interaction-aware target represen-
75 tations. After each of SCISOR’s 12 backbone layers, peptide residues attend to these representations
76 through cross-attention. The resulting features are modulated by a learnable scalar gate initialized to
77 zero, allowing the conditioning pathway to be introduced gradually during fine-tuning while leaving
78 the pretrained model unchanged at initialization. Together, the conditioning modules introduce
79 approximately 15.2M additional parameters (15.0M from cross-attention and 0.19M from the global
80 pathway), increasing the model size from 52.6M parameters. Because all newly introduced projec-
81 tions are zero-initialized, conditioned and unconditioned predictions are identical at initialization,
82 a property we verify before each training run. At inference time, conditioning is optional. Uncon-
83 ditional generation disables both pathways, whereas conditional generation applies a shared target
84 representation across the batch.

85 2.2 Data

86 We combine the PepMLM dataset with Propedia complexes [Chen et al., 2025, Martins et al., 2021],
87 retaining peptide–target pairs with peptides of length ≤ 50 and targets of length ≤ 1024 . The
88 resulting training set contains 52,461 pairs (median peptide length: 11). The PepMLM held-out set
89 contains 293 pairs within the same length range. To prevent data leakage, we removed from the
90 training set all peptide–target pairs and target sequences appearing in either the held-out test set or
91 our binder evaluation benchmarks. We further excluded all complexes used in the crystal-shrinking
92 and *de novo* binder-design evaluations, ensuring that every reported benchmark target and complex
93 was unseen during training.

94 3 Evaluation

95 Our evaluation addresses three questions: (i) does fine-tuning preserve SCISOR’s ability to generate
96 realistic peptide sequences? (ii) Does target conditioning improve peptide shrinking while preserving
97 binding-relevant residues? (iii) Can the model generate plausible *de novo* peptide binders for
98 previously unseen therapeutic targets?

99 3.1 Fitting the distribution of natural sequences

100 **Perplexity** We report a variational upper bound on the negative log-likelihood over the 293 held-out
101 peptide–target pairs. Target conditioning is disabled to evaluate the underlying peptide language
102 model rather than target-specific generation.

103 **Fréchet Protein Distance (FPD)** We generate 1,000 peptide samples using 20 diffusion steps and
104 $K=5$ corrector steps, following the SCISOR protocol. Distributional similarity to natural peptides is
105 measured using FPD in ProtT5 embedding space [Elnaggar et al., 2022]; lower values indicate closer
106 agreement with the distribution of natural peptides.

107 **Monomer foldability** We estimate monomer foldability using OmegaFold with one recycle [Wu
108 et al., 2022], reporting pLDDT for generated peptides and comparing it with the held-out test set.

109 3.2 Shrinking Panel

110 While unconditional generation evaluates peptide realism, binder shrinking evaluates whether target
111 conditioning preserves functionally important residues during sequence compression. We use ten
112 heteromeric peptide–protein crystal complexes (binder lengths 31–50 residues with 7–30 interface
113 residues; Supplementary Material). Each complex is shortened to peptide lengths of 25, 20, 15, and
114 10 residues using three strategies: uniform random deletion, pretrained SCISOR, and IDPepDesign
115 with target conditioning enabled. Five sequences are generated per target length. PepMLM cannot
116 perform sequence shortening and is therefore evaluated only as a *de novo* peptide generator at the
117 corresponding target lengths. For the shortening methods, we measure longest-common-subsequence
118 coverage and the fraction of preserved interface residues relative to the crystal complex [Baron et al.,
119 2025]. Each shortened peptide is subsequently folded together with its target using Boltz [Wohlwend
120 et al., 2024], and the predicted complexes are evaluated using ipTM and ipSAE_{min} [Dunbrack, 2025].
121 Crystal complexes serve as the native reference.

122 3.3 Binder Design against Target Panel

123 We evaluate *de novo* peptide binder generation on 11 of the 15 therapeutic targets from the experi-
124 mental binder meta-analysis of Overath *et al.* [Overath et al., 2025], covering diverse protein classes
125 including receptors, viral proteins, and intracellular protein–protein interaction hubs.

- 126 • All methods generate peptides of fixed length ($L=20$), eliminating peptide length as a
127 confounding factor for interface quality metrics.
- 128 • We sample 20 peptide designs per target and evaluate every generated sequence without
129 pre-filtering or reranking prior to structure prediction [Ko and Wang, 2025].

Table 1: Evaluating the Fine-tuning gain on FPD, perplexity, and pLDDT. Initial test peptides have mean pLDDT 67.2.

Cloud	NLL	PPL	pLDDT	FPD
SCISOR	2.84	17.10	64.8	1.84
IDPepDesign uncond	2.42	11.25	66.0	0.75

- We compare IDPepDesign against PepMLM-650M [Chen et al., 2025] and BoltzGen [Stark et al., 2025]Design.

For each method, we report the mean ipTM and ipSAE_{min} across targets, together with the fraction of designs exceeding established interface-confidence thresholds, averaging across targets rather than reporting best-case examples.

3.3.1 Binder Evaluation metrics

Sequence diversity Sequence diversity is measured as the mean pairwise sequence identity among peptides generated for each target. Lower identity indicates greater diversity.

ipSAE & ipTM Experimental affinity measurements remain the gold standard but are impractical at this scale. We therefore evaluate predicted complexes using two complementary interface confidence metrics. ipTM measures AlphaFold’s interface confidence [Abramson et al., 2024], whereas ipSAE focuses on residue pairs with confident predicted aligned error and better isolates interface quality [Dunbrack, 2025]. A recent meta-analysis of 3,766 experimentally characterized *de novo* binders identified ipSAE_{min} as the strongest single predictor of experimental success among commonly used confidence metrics [Overath et al., 2025]. We therefore report both ipTM and ipSAE_{min} using Boltz [Wohlwend et al., 2024]. These metrics estimate interface confidence rather than binding affinity.

4 Results

4.1 Fine-tuning improves peptide generation

Perplexity Fine-tuning substantially improves sequence likelihood on the held-out peptide–target pairs, reducing perplexity from 17.10 for the pretrained SCISOR model to 11.25 for IDPepDesign (Table 1). This improvement indicates that the model successfully adapts from general protein sequence generation to the distribution of short peptide binders.

Monomer foldability Generated peptides remain structurally plausible after fine-tuning. OmegaFold predicts pLDDT values within approximately two points of the natural peptide distribution (67.2) for all generated sequence sets, with unconditional IDPepDesign being closest. Since binder quality depends primarily on the protein–peptide interface rather than the stability of the isolated peptide, this experiment serves to verify that fine-tuning preserves SCISOR’s ability to generate structurally feasible peptide sequences.

FPD IDPepDesign achieves a lower FPD than the pretrained SCISOR model, indicating that generated peptides more closely match the distribution of natural peptide sequences. At the same time, generation diversity remains unchanged: every sampled sequence is unique, and generated peptide lengths closely match the held-out test distribution (mean ~ 14 residues).

4.2 Interaction-aware peptide shrinking

Figure 1 summarizes peptide shrinking across ten experimentally resolved peptide–protein complexes. Panel A reports longest-common-subsequence coverage of the crystal interface, while Panel B reports the fraction of preserved interface residues. Because PepMLM generates peptides *de novo* rather than shortening existing binders, it is not applicable for the residue-preservation analysis. IDPepDesign consistently preserves more of the crystal interface than the pretrained SCISOR model across all target peptide lengths. More importantly, this improved sequence preservation translates into higher predicted interface quality after structure prediction. Averaged across the ten complexes, IDPepDesign achieves mean ipTM/ipSAE_{min} scores of 0.51/0.042 compared with 0.48/0.037 for SCISOR. At the biologically relevant peptide length of 15 residues, the advantage

Table 2: Mean $\pm$ standard deviation of ipTM and ipSAE_{min} across 10 peptide–protein benchmark targets. Best mean per target/metric is highlighted in and second-best in .

Model	Metric	TnC-TnI	Kallikrein-Hirustasin	MLC1P-Myo2IQ	TBP-TFIIA	YscE-YscG	SGPB-OMTKY3	ExoT-SpcS	UNG-Ugi	TEAD1-YAP	CDK2-p27
IDPepDesign	ipTM	0.78 $\pm$ 0.11	0.67 $\pm$ 0.12	0.67 $\pm$ 0.18	0.51 $\pm$ 0.16	0.22 $\pm$ 0.10	0.59$\pm$0.21	0.55 $\pm$ 0.18	0.25 $\pm$ 0.13	0.39$\pm$0.19	0.45 $\pm$ 0.22
SCISOR		0.75 $\pm$ 0.19	0.64 $\pm$ 0.17	0.64 $\pm$ 0.15	0.44 $\pm$ 0.16	0.21 $\pm$ 0.09	0.52 $\pm$ 0.18	0.52 $\pm$ 0.12	0.26$\pm$0.08	0.32 $\pm$ 0.19	0.48$\pm$0.21
Native		0.78 $\pm$ 0.00	0.48 $\pm$ 0.00	0.61 $\pm$ 0.00	0.49 $\pm$ 0.00	0.20 $\pm$ 0.00	0.48 $\pm$ 0.00	0.90$\pm$0.00	0.14 $\pm$ 0.00	0.19 $\pm$ 0.00	0.15 $\pm$ 0.00
PepMLM		0.76 $\pm$ 0.15	0.70$\pm$0.10	0.72$\pm$0.13	0.54$\pm$0.18	0.27$\pm$0.16	0.52 $\pm$ 0.19	0.59 $\pm$ 0.17	0.25$\pm$0.08	0.35$\pm$0.22	0.56$\pm$0.19
Random		0.80$\pm$0.09	0.67 $\pm$ 0.15	0.61 $\pm$ 0.12	0.50 $\pm$ 0.12	0.18 $\pm$ 0.06	0.49 $\pm$ 0.20	0.51 $\pm$ 0.19	0.26$\pm$0.09	0.33 $\pm$ 0.17	0.39 $\pm$ 0.17
IDPepDesign	ipSAE _{min}	0.12 $\pm$ 0.10	0.07 $\pm$ 0.06	0.08 $\pm$ 0.09	0.02 $\pm$ 0.03	0.00 $\pm$ 0.01	0.07$\pm$0.09	0.04 $\pm$ 0.04	0.00$\pm$0.01	0.01 $\pm$ 0.02	0.01 $\pm$ 0.03
SCISOR		0.12$\pm$0.14	0.08 $\pm$ 0.09	0.07 $\pm$ 0.09	0.01 $\pm$ 0.01	0.00 $\pm$ 0.00	0.04 $\pm$ 0.08	0.03 $\pm$ 0.03	0.00 $\pm$ 0.00	0.01 $\pm$ 0.01	0.01 $\pm$ 0.03
Native		0.07 $\pm$ 0.00	0.02 $\pm$ 0.00	0.04 $\pm$ 0.00	0.01 $\pm$ 0.00	0.00 $\pm$ 0.00	0.01 $\pm$ 0.00	0.51$\pm$0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
PepMLM		0.12 $\pm$ 0.12	0.07 $\pm$ 0.04	0.10$\pm$0.09	0.02$\pm$0.02	0.01$\pm$0.02	0.04 $\pm$ 0.05	0.05 $\pm$ 0.06	0.00 $\pm$ 0.00	0.01$\pm$0.03	0.03$\pm$0.09
Random		0.12$\pm$0.10	0.09$\pm$0.09	0.05 $\pm$ 0.04	0.01 $\pm$ 0.01	0.00 $\pm$ 0.00	0.03 $\pm$ 0.03	0.04 $\pm$ 0.05	0.00 $\pm$ 0.00	0.00 $\pm$ 0.01	0.00 $\pm$ 0.01

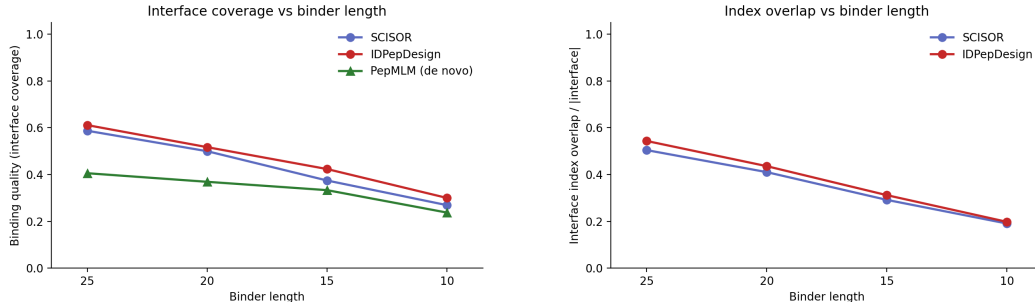


Figure 1: Shrinking ten peptide–protein complexes versus binder length (mean over complexes). **(Left)** Coverage of the crystal interface subsequence. IDPepDesign sits above SCISOR; PepMLM is de novo and lower. **(Right)** Fraction of preserved crystal interface indices. PepMLM is omitted because it does not delete.

173 becomes more pronounced (0.54/0.066 versus 0.50/0.042). Although PepMLM achieves slightly
174 higher mean ipTM, it is unconstrained by the original binder sequence and therefore solves a different
175 design problem. These results indicate that target conditioning enables SCISOR to preferentially
176 retain interaction-relevant residues during sequence compression while producing higher-confidence
177 predicted interfaces (Table 2).

178 4.3 Target-conditioned peptide binder design

179 Table 3 and Figure 4 summarize *de novo* binder generation for 11 therapeutic targets. All methods
180 generate fully diverse peptide sets with no duplicate sequences. However, IDPepDesign consistently
181 exhibits the lowest mean pairwise sequence identity, indicating substantially greater sequence diversity
182 than PepMLM while maintaining comparable predicted interface quality. Despite containing only
183 approximately 53M parameters, IDPepDesign achieves interface confidence comparable to PepMLM
184 (650M parameters) and BoltzGen Design ($\sim$ 500M parameters). This demonstrates that explicitly
185 conditioning an insertion–deletion diffusion model on target sequences provides a highly parameter-
186 efficient strategy for peptide binder generation. Importantly, the increased diversity suggests that
187 IDPepDesign explores a broader region of peptide sequence space without compromising predicted
188 binding quality, making it a promising framework for discovering diverse candidate binders.

189 5 Conclusion

190 We presented **IDPepDesign**, a target-conditioned extension of SCISOR for sequence-only pep-
191 tide binder design. By augmenting an insertion–deletion diffusion model with lightweight target-
192 conditioning pathways, IDPepDesign generates realistic peptides, preserves the pretrained generative
193 capabilities of SCISOR, and achieves competitive peptide binder design despite using substantially
194 fewer parameters than existing methods. Our results further suggest that insertion–deletion diffusion
195 provides an effective inductive bias for identifying and preserving interaction-relevant residues dur-
196 ing peptide binder generation and sequence compression. Future work will focus on experimental
197 validation and scaling the approach to larger peptide–protein interaction datasets.

Table 3: Mean $\pm$ standard deviation of ipTM, ipSAE_{min}, and mean sequence identity on the 11 targets. Best mean per target/metric is highlighted in and second-best in . Lower mean identity indicates higher sequence diversity.

Model	Metric	FGFR2	IL10Ra	IL2Ra	IL7Ra	InsulinR	LTK	MDM2	PD-L1	SARS-CoV-2 RBD	TrkA	VirB8
IDPepDesign	ipTM	0.72$\pm$0.12	0.74$\pm$0.18	0.33 $\pm$ 0.19	0.56 $\pm$ 0.16	0.78 $\pm$ 0.10	0.25 $\pm$ 0.07	0.77 $\pm$ 0.11	0.88$\pm$0.04	0.61$\pm$0.14	0.66 $\pm$ 0.15	0.50 $\pm$ 0.23
	ipSAE _{min}	0.12$\pm$0.11	0.13 $\pm$ 0.13	0.01$\pm$0.03	0.03 $\pm$ 0.03	0.16 $\pm$ 0.12	0.00 $\pm$ 0.00	0.19 $\pm$ 0.13	0.31$\pm$0.11	0.03 $\pm$ 0.04	0.07 $\pm$ 0.05	0.05 $\pm$ 0.11
	Mean identity	0.06$\pm$0.01	0.06 $\pm$ 0.01	0.05 $\pm$ 0.01	0.06$\pm$0.01	0.06$\pm$0.01	0.07$\pm$0.01	0.06$\pm$0.02	0.06$\pm$0.01	0.09 $\pm$ 0.03	0.06$\pm$0.01	0.06$\pm$0.01
Boltzgen	ipTM	0.68 $\pm$ 0.14	0.71 $\pm$ 0.18	0.31 $\pm$ 0.10	0.59 $\pm$ 0.18	0.80 $\pm$ 0.15	0.29$\pm$0.13	0.91$\pm$0.05	0.86 $\pm$ 0.06	0.64$\pm$0.19	0.74$\pm$0.14	0.68$\pm$0.21
	ipSAE _{min}	0.09 $\pm$ 0.11	0.09 $\pm$ 0.09	0.00 $\pm$ 0.00	0.05$\pm$0.05	0.23 $\pm$ 0.16	0.00$\pm$0.01	0.49$\pm$0.15	0.28 $\pm$ 0.16	0.04$\pm$0.06	0.13$\pm$0.11	0.11$\pm$0.13
	Mean identity	0.07 $\pm$ 0.01	0.07 $\pm$ 0.02	0.07 $\pm$ 0.02	0.07 $\pm$ 0.01	0.06 $\pm$ 0.01	0.18 $\pm$ 0.05	0.08 $\pm$ 0.02	0.07 $\pm$ 0.02	0.06$\pm$0.02	0.08 $\pm$ 0.01	0.07 $\pm$ 0.02
PepMLM	ipTM	0.52 $\pm$ 0.18	0.71 $\pm$ 0.16	0.45$\pm$0.16	0.60$\pm$0.14	0.86$\pm$0.07	0.28 $\pm$ 0.10	0.88 $\pm$ 0.05	0.92$\pm$0.03	0.57 $\pm$ 0.16	0.70 $\pm$ 0.10	0.49 $\pm$ 0.19
	ipSAE _{min}	0.04 $\pm$ 0.05	0.08 $\pm$ 0.10	0.01 $\pm$ 0.01	0.04 $\pm$ 0.04	0.29$\pm$0.12	0.00 $\pm$ 0.00	0.39 $\pm$ 0.13	0.42$\pm$0.11	0.02 $\pm$ 0.04	0.10 $\pm$ 0.11	0.02 $\pm$ 0.03
	Mean identity	0.39 $\pm$ 0.03	0.37 $\pm$ 0.03	0.39 $\pm$ 0.03	0.36 $\pm$ 0.03	0.41 $\pm$ 0.04	0.39 $\pm$ 0.04	0.43 $\pm$ 0.04	0.40 $\pm$ 0.04	0.41 $\pm$ 0.05	0.34 $\pm$ 0.03	0.37 $\pm$ 0.04

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 Židek, 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.
- Sarah Alamdari, Nitya Thakkar, Rianne van den Berg, Alex X Lu, Nicolo Fusi, Ava P Amini, and Kevin K Yang. Protein generation with evolutionary diffusion: sequence is all you need. *bioRxiv*, 2023.
- Ethan Baron, Alan Nawzad Amin, Ruben Weitzman, Debora Susan Marks, and Andrew Gordon Wilson. A diffusion model to shrink proteins while maintaining their function. *arXiv preprint arXiv:2511.07390*, 2025.
- Leo Tianlai Chen, Zachary Quinn, Madeleine Dumas, Christina Peng, Lauren Hong, Moises Lopez-Gonzalez, Alexander Mestre, Rio Watson, Sophia Vincoff, Lin Zhao, et al. Target sequence-conditioned design of peptide binders using masked language modeling. *Nature Biotechnology*, 2025.
- 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, et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022.
- Roland L Dunbrack. ipSAE: A tool for quantifying interprotein contacts from predicted complexes. *bioRxiv*, 2025.
- Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, et al. ProtTrans: Toward understanding the language of life through self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7112–7127, 2022.
- Young Su Ko and Wei Wang. forge: sequence-based binder design with latent flow matching. In *Machine Learning for Structural Biology Workshop*, 2025.
- Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.
- Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *International Conference on Learning Representations*, 2019.

- 236 Pedro M Martins, Lucianna H Santos, Diego Mariano, Felipe C Queiroz, Luana L Bastos, Isabela
237 de S Gomes, Pedro H Fischer, Raquel E Rocha, Sabrina A Silveira, Luiz Henrique V de Lima,
238 et al. Propedia: a database for protein–peptide identification in PDB and construction of structural
239 networks. *BMC Genomics*, 22:1–15, 2021.
- 240 Max D. Overath, Andreas S. H. Rygaard, Christian P. Jacobsen, Valentas Brasas, Oliver Morell, Pietro
241 Sormanni, and Timothy P. Jenkins. Predicting experimental success in de novo binder design: A
242 meta-analysis of 3,766 experimentally characterised binders. August 2025.
- 243 Martin Pacesa, Lennart Nickel, Christian Schellhaas, Joseph Schmidt, Ekaterina Pyatova, Srajan
244 Kaplanis, et al. One-shot design of functional protein binders with BindCraft. *Nature*, 2025.
- 245 Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy O’Donnell, Anton
246 Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, Mateo Reveiz, Roman Bushuiev, Tally Portnoi,
247 Tomáš Pluskal, Josef Sivic, Karsten Kreis, Arash Vahdat, Shamayeeta Ray, Jonathan T. Goldstein,
248 Andrew Savinov, Jacob A. Hambalek, Anshika Gupta, Diego A. Taquiri-Diaz, Yaotian Zhang,
249 Samuel J. Snyder, A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Haoyu Fan, Ethel
250 Tackie-Yarboi, Dylan Boselli, Lee Schnaider, Chang C. Liu, Gene-Wei Li, Denes Hnisz, David M.
251 Sabatini, William F. DeGrado, Jeremy Wohlwend, Gabriele Corso, Regina Barzilay, and Tommi
252 Jaakkola. Boltzgen: Toward universal binder design. *bioRxiv*, 2025. doi: 10.1101/2025.11.20.
253 689494.
- 254 Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach,
255 Woody Ahern, Alexander J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of
256 protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, 2023.
- 257 Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Noah Getz, Mateo Reveiz, Ken Leidal, Wojtek
258 Swiderski, Liam Atkinson, Tally Portnoi, Itamar Chinn, et al. Boltz-1: Democratizing Biomolecular
259 interaction modeling. *bioRxiv*, 2024.
- 260 Ruidong Wu, Fan Ding, Rui Wang, Rui Shen, Xiwen Zhang, Shitong Luo, Chenpeng Su, Zuofan Wu,
261 Qi Xie, Bonnie Berger, et al. High-resolution de novo structure prediction from primary sequence.
262 *bioRxiv*, 2022.

263 A Training

264 Figure 2 shows the IDPepDesign denoising network. A noisy binder sequence is embedded and
265 processed by N DiT blocks. Diffusion time t and the pooled target representation are fused into
266 a global conditioning vector that produces AdaLN scale/shift parameters. Residue-level target
267 embeddings enter each block through cross-attention. The head outputs deletion logits over binder
268 positions.

269 A.1 Optimization

270 We fine-tune for 5 epochs with AdamW [Loshchilov and Hutter, 2019], learning rate 10^{-4} , linear
271 warmup of 1,000 optimizer steps, gradient clipping 1.0, and no weight decay. Preprocessed batches
272 have size 8; we accumulate 32 steps for an effective batch of 256, matching SCISOR. That is $\sim 1,023$
273 optimizer steps in total. New parameters start at the zero-gate; the pretrained peptide ESM backbone
274 and FiLM weights are not frozen. The frozen target ESM-2 is used only to precompute embeddings
275 and is not in the training graph.

276 A.2 Hardware and resources

277 Fine-tuning, ELBO evaluation, $K=5$ sampling, and OmegaFold each run on one NVIDIA A100.

278 B Hyperparameters

279 Table 4 lists the architecture and optimization settings used for all experiments. FiLM width is
280 inherited from the SCISOR checkpoint and is not changed during fine-tuning. Cross-attention is
281 applied after every backbone layer.

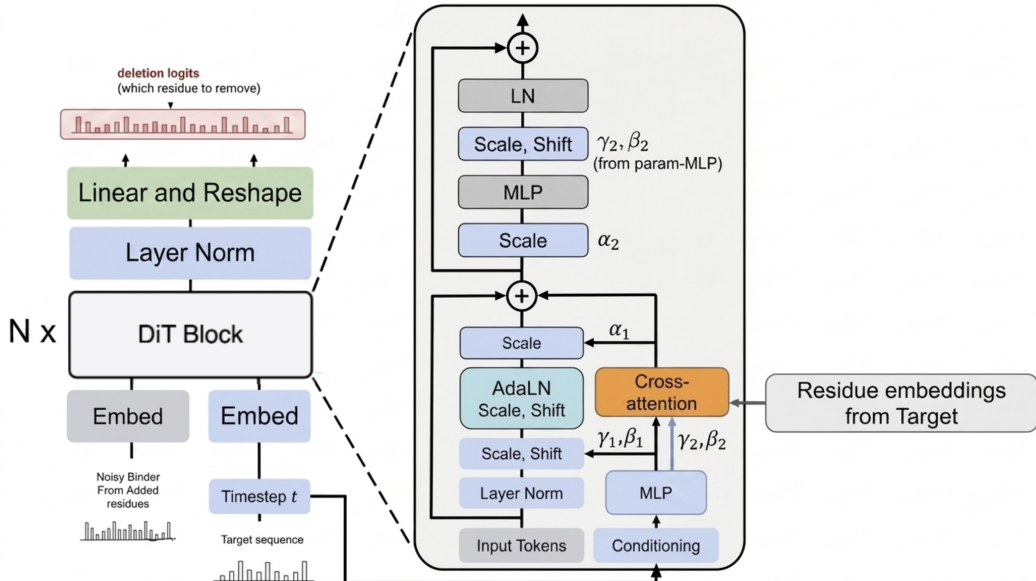


Figure 2: IDPepDesign architecture. **(Left)** The denoising network takes a noisy binder (insertions added by the forward process), diffusion time t , and the target sequence. After N DiT blocks, a LayerNorm and linear head produce deletion logits indicating which residues to remove. **(Right)** Each DiT block applies AdaLN using parameters predicted from the global conditioning vector (time + pooled target), then peptide-to-target cross-attention over residue embeddings of the target, followed by an MLP with residual connections. Newly introduced projections are zero-initialized so that the pretrained SCISOR model is recovered at the start of fine-tuning.

Table 4: Training and architecture. FiLM width is pinned by the SCISOR checkpoint.

Item	Value
Init checkpoint	SCISOR_U90_S
Peptide backbone	ESM-2 35M (esm2_t12_35M_UR50D), 12 layers
Frozen target encoder	ESM-2 35M, residue dim 480
FiLM width d	256
Target encoder layers	2
Cross-attention	after every backbone layer, 20 heads
New parameters	$\sim 15.2\text{M} / 52.6\text{M}$ total
$t_{\max}, \alpha\text{-rate}$	0.9, 1.1
Train pairs / test pairs	52,461 / 293
Peptide length (train)	min 2, median 11, max 50
Optimizer	AdamW, lr 10^{-4} , warmup 1,000 steps, clip 1.0
Effective batch	$8 \times 32 = 256$
Epochs / optimizer steps	5 / $\sim 1,023$

282 C Shrinking Panel

283 Table 5 lists the ten heteromeric peptide–protein crystals used in Section 3.2. Complexes were
 284 chosen so that the binder is long enough to shorten into the peptide window while remaining a
 285 peptide–protein pair rather than a homomer. Binders are 31–50 residues, with 7–30 interface residues
 286 (heavy atoms within 4 Å of the target). Each binder is shortened to lengths 25, 20, 15, and 10. Target
 287 and binder columns are the crystal chain IDs used for shrinking.

Table 5: Shrinking panel. Heteromeric peptide–protein crystals; the binder is shortened into the peptide window. Interface size is the number of binder residues with a heavy atom within 4 Å of the target.

PDB	Complex	Target	Binder	L	iface
1A2X	TnC–TnI	A	B	31	19
3SGB	SGPB–OMTKY3	E	I	50	9
1HIA	kallikrein–hirustasin	B	I	48	9
4RE1	TEAD1–YAP	B	C	50	30
6ATH	CDK2–p27	B	C	50	16
2P58	YscE–YscG	C	A	50	14
1N2D	MLC1P–Myo2IQ	A	C	48	22
4JMF	ExoT–SpcS	C	A	50	16
1NH2	TBP–TFIIA	A	C	50	7
4L5N	UNG–Ugi	A	C	50	15

D Sampling defaults

Unless noted, generation uses 20 diffusion steps, temperature 1, $r=1$, and $K=5$ corrector steps. Shrinking uses the same temperature and a deletion budget $S = L_{\text{binder}} - L$, with $L \in \{25, 20, 15, 10\}$ (Figure 1). PepMLM is run at the same output length as IDPepDesign and SCISOR.

E Amino-acid composition

Fine-tuning should move generated sequences off the UniRef protein composition and onto the peptide training distribution. Figure 3 compares residue frequencies of SCISOR samples, IDPepDesign samples, and natural test peptides. IDPepDesign tracks the peptide set more closely than SCISOR, consistent with the FPD drop from 1.84 to 0.75.

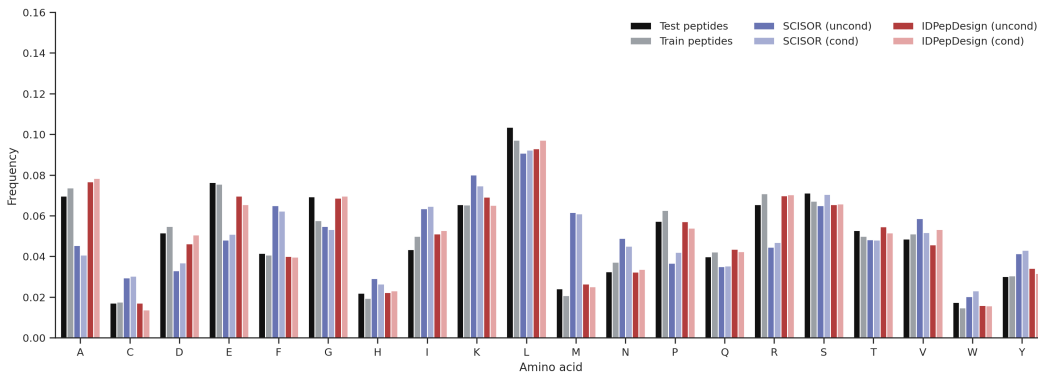


Figure 3: Amino-acid composition of generated clouds versus natural peptides. Fine-tuning shifts SCISOR toward the peptide residue frequencies; IDPepDesign with and without a target remain close to each other.

F FPD and pLDDT

Fréchet protein distance (FPD). We calculate FPD of generated sequences to natural peptides using ProtT5 embeddings [Elnaggar et al., 2022] as

$$\|\mu_{\text{natural}} - \mu_{\text{sample}}\|^2 + \text{tr}(\Sigma_{\text{natural}} + \Sigma_{\text{sample}} - 2(\Sigma_{\text{natural}}\Sigma_{\text{sample}})^{1/2})$$

where μ . and Σ . are empirical means and covariances of the embeddings. pLDDT. We calculate pLDDT scores using OmegaFold [Wu et al., 2022]. For computational efficiency, we use only 1 cycle per sample. This results in lower overall pLDDT scores than the recommended default of 10 cycles.

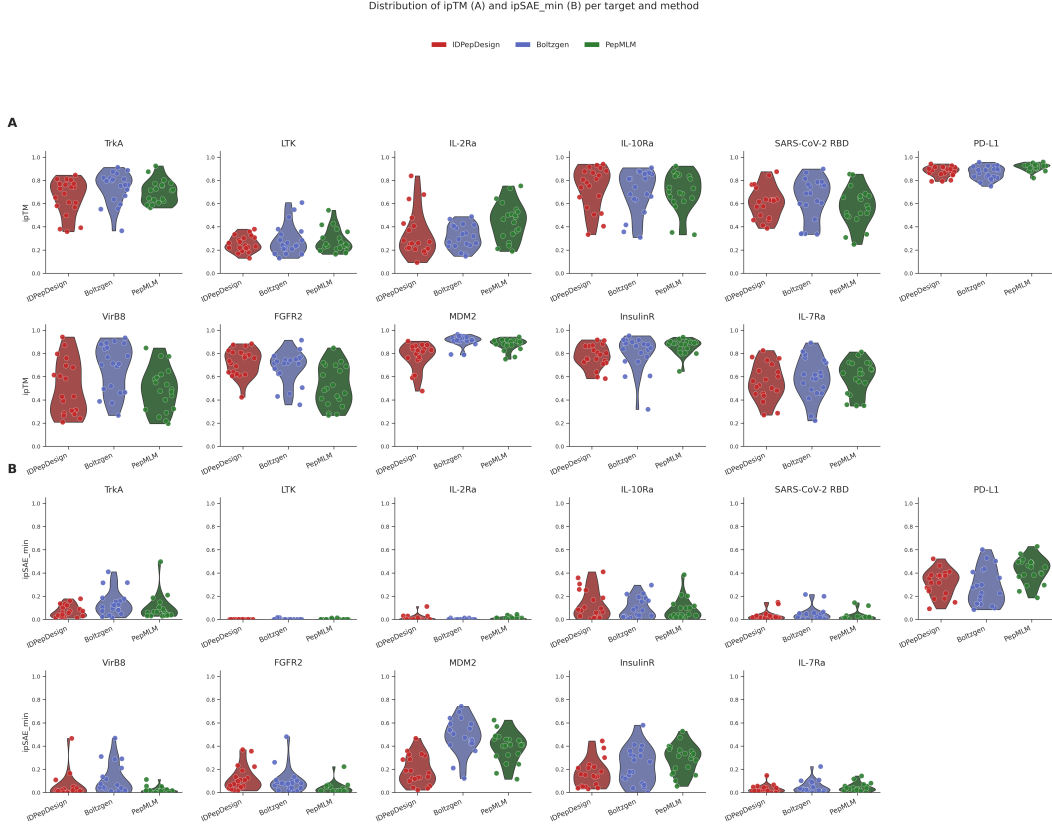


Figure 4: Distribution of ipTM (A) and ipSAE_{min} (B) on the 11-target panel. Each point is one unfiltered 20-mer folded with the target. IDPepDesign matches PepMLM and BoltzGen on typical interface confidence, with target-dependent spread (high on PD-L1, FGFR2, MDM2; near zero on LTK).

G Binder-design panel

Figure 4 shows the per-design distribution of ipTM and ipSAE_{min} on the 11-target panel from Section 3.3. Each point is one unfiltered 20-mer folded with the target; no sequences are ranked or discarded before structure prediction. Spread is target-dependent: several receptors yield high interface confidence for all methods, while LTK remains near zero.