PepSeek: RECEPTOR-CONDITIONED REPRESENTATION LEARNING FOR PEPTIDE RECOGNITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Peptide discovery requires ranking vast peptide–receptor spaces before expensive structural and experimental evaluation. We introduce PepSeek, a coarse-to-fine receptor-conditioned workflow that couples hierarchical molecular graph learning with amortized complex prediction. Its ranking stage embeds peptide atom–bond graphs and receptor residue–contact graphs, exchanges information between peptide motifs and learned receptor clusters, and adapts only task-relevant terminal parameters. On P1, a leakage-controlled cold-chemical-series benchmark spanning five activity endpoints, the metric-specific diagnostic envelope achieved the best evaluated RMSE and Spearman correlation for every endpoint; relative to the strongest non-PepSeek comparator, RMSE decreased by 1.1–23.0% and Spearman ρ increased by 0.064–0.230. Across 16 GPCRs, receptor-specific adaptation improved ranking for 13 receptors. Ranking enriched high-confidence MC4R complexes 6.22-fold among the top 25 peptides. AF-turbo then reused receptor context to reduce median complex-prediction time 7.26-fold while preserving mean DockQ. Cellular assays identified seven concentration-dependent responses among 32 evaluable peptides. These results show that receptor-conditioned ranking and cached structural refinement can allocate finite compute and assay budgets to peptide candidates with convergent statistical, structural and functional evidence.

1 INTRODUCTION

Proteome-derived peptide discovery is a coupled representation and resource-allocation problem. Candidate libraries contain related chemotypes, assay endpoints differ in meaning and noise, and only a small fraction of peptide–receptor pairs can receive repeated structure prediction or experimental testing. A useful system must therefore learn transferable molecular features, preserve receptor-specific interaction geometry and expose a calibrated ranking that can contract the search space without conflating computational confidence with biological activity.

PepSeek addresses these requirements with two coordinated stages (Figure 1). The first represents a peptide as an atom–bond graph and a receptor as a residue graph containing sequence and predicted-contact edges. Hierarchical message passing couples peptide motifs to learned receptor clusters, producing a receptor-conditioned interaction fingerprint that can be adapted with a small trainable subset. The second stage applies AF-turbo to the resulting shortlist, reusing receptor-side alignment and template context across peptides while retaining matched seeded complex prediction and explicit interface gates.

The study contributes a mathematically specified coarse-to-fine architecture, the P1 cold-chemical-series benchmark for five heterogeneous activity endpoints, and an end-to-end evaluation spanning receptor ranking, structural enrichment, compute efficiency and cellular response. We test receptor-specific ranking across 16 GPCRs, MC4R enrichment within a shared 7,561-peptide pool, runtime and interface accuracy across 59 paired complexes, and cellular responses among synthesized candidates. This design distinguishes three questions that are often merged: whether a model extrapolates to unseen chemical series, whether ranking concentrates structurally credible complexes, and whether selected peptides show measurable receptor activity.

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.

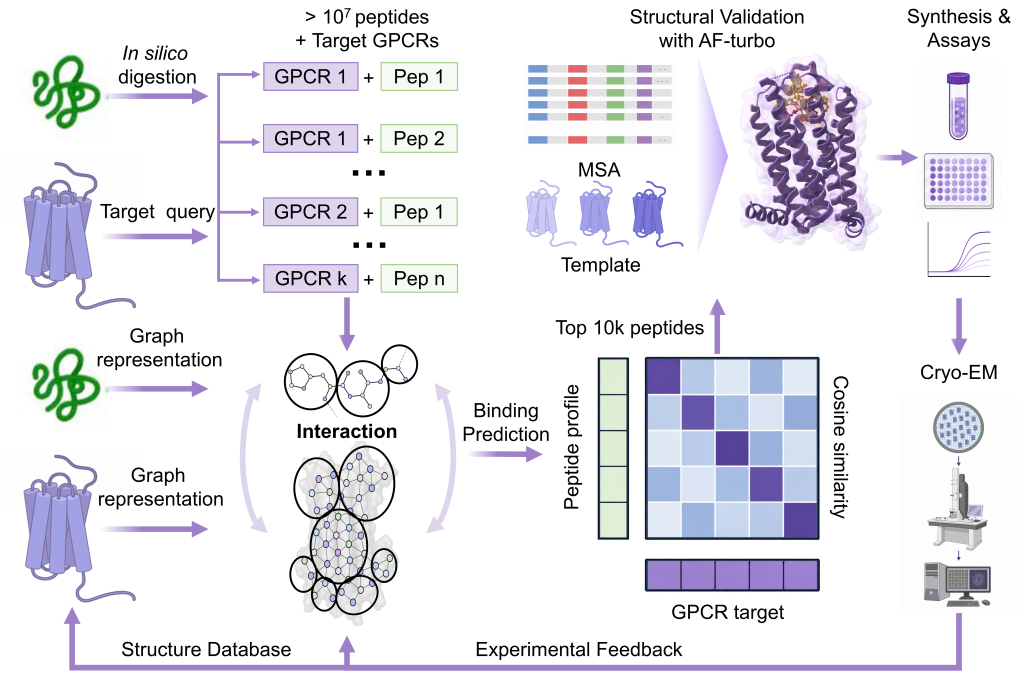

Figure 1. PepSeek workflow for GPCR-targeted peptide discovery. Food-protein digestion generates candidate peptides for receptor-conditioned ranking. Peptide atom–bond graphs and receptor residue graphs are processed by an interaction network. High-ranking candidates undergo AF-turbo complex prediction, followed by synthesis and functional characterization. Receptor sequence, alignment and template context are reused across candidates. The schematic illustrates budgeted selection and information flow; the evaluated cohorts are defined in the text.

2 Related Work

Protein and peptide representation learning. Protein language models learn transferable residue representations from large sequence corpora (Rao et al., 2019; Rives et al., 2021; Elnaggar et al., 2022; Lin et al., 2023). Graph neural networks provide complementary relational and geometric inductive biases (Kipf & Welling, 2017; Veličković et al., 2018; Satorras et al., 2021; Jing et al., 2021). GearNet uses relation-specific residue edges and edge message passing (Zhang et al., 2023), whereas LM–GVP combines pretrained sequence features with geometric reasoning (Wang et al., 2022). In affinity prediction, GraphDTA represents compounds as graphs, MGraphDTA deepens multiscale molecular encoding, and DeepPurpose supplies modular compound and protein encoders (Nguyen et al., 2021; Yang et al., 2022; Huang et al., 2020). PepSeek instead makes the peptide–receptor interaction itself hierarchical: sequence features supply biochemical context, molecular edges preserve peptide chemistry and receptor contacts, and cross-entity attention learns which motif–pocket relations determine the score.

Few-shot adaptation for biological prediction. Meta-learning and parameter-efficient adaptation reuse shared features when task-specific supervision is scarce (Finn et al., 2017; Snell et al., 2017). Protein-language-model meta-learning similarly enables adaptation with limited labels (Beck et al., 2025), and representation-learning theory explains how a shared representation can reduce target-task sample complexity (Du et al., 2021). PepSeek preserves the common molecular encoders and adapts the terminal interaction layers that determine how peptide features are interpreted for a receptor, separating receptor conditioning from wholesale retraining.

Structure prediction and virtual screening. Modern predictors provide accurate protein structures and increasingly capable complex hypotheses (Jumper et al., 2021; Abramson et al., 2024), and folding networks can support peptide–protein docking (Tsaban et al., 2022). Structure-conditioned

design methods include ProteinMPNN, RFdiffusion, backbone diffusion and joint sequence–structure models (Dauparas et al., 2022; Watson et al., 2023; Trippe et al., 2023; Shi et al., 2023); DiffDock highlights confidence and out-of-distribution behavior in docking (Corso et al., 2023), while surface models show the value of interface geometry (Gainza et al., 2020; Tubiana et al., 2022). GPCRdb provides receptor annotations and structures (Herrera et al., 2025), and MC4R structures resolve the binding pocket and calcium cofactor (Yu et al., 2020). PepSeek applies structure prediction after ranking and amortizes receptor-side preprocessing across shortlisted peptides.

3 Problem Setting

3.1 Receptor-specific prediction

Let $\mathcal{P}$ denote a peptide library and $\mathcal{T}$ a set of receptor tasks. For receptor t, the labelled set $\mathcal{D}_t = \{(p_i, y_i)\}_{i=1}^{n_t}$ contains peptide sequences and assay-specific measurements. A score $s(p,t)$ orders peptides within a receptor and assay endpoint. Adaptation uses training observations; evaluation uses held-out observations.

The structural corpus contains 47,032 chain-pair records (30,028 unique sequence pairs), of which 18,281 records (11,831 pairs) passed quality screening; the receptor-ranking benchmark contains 493 labelled observations across 16 GPCRs.

For the 16-receptor analysis, we used five-fold within-receptor cross-validation with labels sorted and assigned cyclically to folds. Each training fold contained 16–44 observations. Receptor-level Spearman correlations were computed from pooled out-of-fold predictions.

3.2 Budgeted candidate selection

Let $\mathcal{S}^{(0)} = \mathcal{P} \times \mathcal{T}$ be the candidate interaction space. Stage ℓ applies a score s_ℓ, an acceptance gate g_ℓ and a candidate budget K_ℓ:

$$\mathcal{S}^{(\ell+1)} = \mathrm{TopK}_{K_\ell}\Big(\{x \in \mathcal{S}^{(\ell)} : g_\ell(x) = 1\}; s_\ell\Big). \tag{1}$$

For $N_\ell = |\mathcal{S}^{(\ell)}|$ candidates and per-candidate cost c_ℓ, the total evaluation cost is

$$C_{\mathrm{total}} = \sum_{\ell=0}^{L-1} N_\ell c_\ell \le B, \tag{2}$$

where B is the available budget. Ranking concentrates structural and experimental evaluation on selected candidates.

The food-derived library comprises 509,425 unique peptides, including 75,767 of at most 25 residues (Figure S1). A shared set of 7,561 peptides was scored against MC1R, MC4R, NK2R, GLP1R and F2R, with structural enrichment evaluated against MC4R. Of 40 synthesized candidates, 32 were evaluable and seven elicited cellular responses.

3.3 Evaluation quantities

Ranking performance measures within-receptor agreement with quantitative labels. Structural enrichment measures the prevalence of high-confidence complexes among high-ranking candidates. Cellular assays characterize the responses of synthesized peptides. For background pool $\mathcal{B}_t$ and structural criterion H,

$$\mathrm{Enr@}K = \frac{K^{-1} \sum_{p \in \mathrm{TopK}_s(t)} H(p,t)}{|\mathcal{B}_t|^{-1} \sum_{p \in \mathcal{B}_t} H(p,t)}. \tag{3}$$

Structural positives were defined by receptor–peptide ipTM ≥ 0.60. Seven of the top 50 peptides met this threshold, compared with 243 of 7,561 candidates overall. Top-50 enrichment was 4.36-fold, with a conditional bootstrap 95% confidence interval of 1.85–7.47.

4 METHOD

4.1 MOLECULAR GRAPH REPRESENTATIONS

For peptide p and receptor t, the scorer receives an atom–bond graph $G_p = (V_p, E_p^{\text{bond}}, M_p)$ and a residue graph $G_t = (V_t, E_t^{\text{seq}} \cup E_t^{\text{contact}})$ (Koh et al., 2024). M_p maps atoms to junction-tree motifs. A peptide atom a carries its element identity z_a and a 43-dimensional physicochemical descriptor x_a; a receptor residue r carries a 33-dimensional descriptor x_r and a 1,280-dimensional sequence embedding e_r (Lin et al., 2023). Receptor edges connect sequence offsets $\pm 1, \pm 2$ and predicted contacts of probability at least 0.5; radial basis functions encode contact strength. Separate projections place both entities in the same 200-dimensional hidden space:

$$h_a^{(0)} = E_z(z_a) + f_a(x_a), \qquad h_r^{(0)} = f_r(x_r) + f_e(e_r) \in \mathbb{R}^{200}. \tag{4}$$

Thus the peptide representation preserves covalent chemistry, whereas the receptor representation combines local sequence continuity, predicted tertiary contacts and language-model context.

4.2 HIERARCHICAL GRAPH INTERACTION

At interaction stage ℓ, principal-neighbourhood aggregation (PNA) summarizes messages by mean, minimum, maximum and standard deviation, then applies identity, amplification and attenuation degree scalers (Corso et al., 2020). Writing $m_{uv}^{(\ell)} = \varphi_\ell(h_u^{(\ell)}, h_v^{(\ell)}, e_{uv})$, the update is

$$\begin{aligned} \mathcal{A}_v^{(\ell)} &= \mathbf{s}_{d_v} \odot \Big\|_{A \in \{\mu, \min, \max, \sigma\}} A_{u \in \mathcal{N}(v)} m_{uv}^{(\ell)}, \\ h_v^{(\ell+1)} &= \mathrm{GraphNorm}\Big[h_v^{(\ell)} + \psi_\ell\Big(h_v^{(\ell)}, \mathcal{A}_v^{(\ell)}\Big)\Big]. \end{aligned} \tag{5}$$

where $\mathbf{s}_{d_v}$ depends on node degree and $\|$ denotes concatenation. Degree scaling reduces the tendency of common high-degree receptor residues or peptide atoms to dominate the representation.

The model then resolves interactions at two matched levels. Junction-tree pooling produces peptide motifs $C_p^{(\ell)} = M_p^\top H_p^{(\ell)}$ (Jin et al., 2018). A learned assignment $S_t^{(\ell)} = \mathrm{softmax}[\mathrm{GCN}_\ell(H_t^{(\ell)}, A_t)]$ groups receptor residues into 5, 10 and 20 clusters across the three stages, with min-cut regularization encouraging compact, separated modules (Bianchi et al., 2020). For attention head h, peptide motif m and receptor cluster k interact through

$$\alpha_{mk}^{(\ell,h)} = \mathrm{softmax}_k\left(\frac{(W_{q,h} c_{p,m}^{(\ell)})^\top (W_{k,h} c_{t,k}^{(\ell)})}{\sqrt{d_h}}\right), \qquad \widetilde{c}_{p,m}^{(\ell,h)} = \sum_k \alpha_{mk}^{(\ell,h)} W_{v,h} c_{t,k}^{(\ell)}. \tag{6}$$

Five attention heads exchange information between peptide motifs and receptor clusters; the same correspondence weights carry peptide values back to receptor clusters. Motif and cluster updates are projected to atoms and residues through M_p and $S_t^{(\ell)}$, combined with residual states and graph-normalized (Cai et al., 2021). Accumulated stage-wise attention produces final atom and residue weights β^p and β^t:

$$\begin{aligned} u_p &= \sum_{a \in V_p} \beta_a^p W_p h_a, \qquad & u_t &= \sum_{r \in V_t} \beta_r^t W_t h_r, \\ z_{p,t} &= [u_p \| u_t] \in \mathbb{R}^{400}, \quad & s_{\theta,\phi}(p,t) &= \mathrm{MLP}_{400 \to 200 \to 1}(z_{p,t}). \end{aligned} \tag{7}$$

This construction makes the prediction explicitly receptor-conditioned: the peptide contribution depends on the receptor clusters to which its motifs attend, and conversely the receptor summary depends on peptide-facing clusters. Input encoders and final pooling projections remain fixed during adaptation; trainable parameters ϕ comprise the regression head and, for selected tasks, the terminal interaction module.

4.3 ACTIVITY CORPUS, QUALITY CONTROL AND P1 PARTITION

Frozen ChEMBL and GPCRdb-derived snapshots supplied structures, receptor sequences, assay context and provenance. Deduplication yielded 32,258 activity groups from 39,076 source records; task mapping retained 31,671 groups for pKi, pKd, binding pIC50, functional pIC50 and pEC50.

Quality control reconciled molecular and receptor identities, graph construction, endpoint semantics and censoring, and exported immutable identifiers with SHA256 manifests.

P1 denotes the cold-chemical-series protocol. Records connected through a shared chemical series, peptide–receptor pair or original measurement were assigned to one component by a deterministic, label-value-blind procedure. Components were allocated once toward 70/15/15 train/validation/test targets, and identities with documented historical evaluation exposure were retained in training. This yielded 28,686 P1 records across the five tasks (22,727/3,315/2,644 train/validation/test; Table S5). Component-level assertions verified that chemical series, pairs and measurements did not cross partitions. Unlike a random record split, P1 prevents near-duplicate chemotypes and repeated measurements from inflating performance, so its test partition measures transfer to unseen chemical series within each endpoint. The indivisibility of large components can make record counts deviate from nominal proportions, but preserving chemical-series integrity is the property required for prospective peptide prioritization.

4.4 Adaptation objective and optimization

For observation i, the admissible interval $[l_i, u_i]$ equals the reported value for exact measurements and retains the reported bound for censored measurements. With prediction s_i, the interval distance and Huber term are

$$d_i = \max(l_i - s_i, 0) + \max(s_i - u_i, 0), \qquad \mathcal{L}_{\text{int}} = \frac{1}{N}\sum_i h_1(d_i), \tag{8}$$

where $h_1(d) = d^2/2$ for $d \le 1$ and $h_1(d) = d - 1/2$ otherwise. Reliable within-group exact-label pairs with activity separation ≥ 0.5 contribute

$$\mathcal{L}_{\text{rank}} = \frac{1}{|\mathcal{R}|}\sum_{(i,j)\in\mathcal{R}} \operatorname{softplus}\left[-\operatorname{sign}(y_i - y_j)\frac{s_i - s_j}{0.5}\right], \quad \mathcal{L} = \mathcal{L}_{\text{int}} + \lambda_r\mathcal{L}_{\text{rank}} + \lambda_a\frac{\|\phi - \phi_0\|_2^2}{|\phi|}. \tag{9}$$

Ranking gradients were capped at 25% of the supervised loss per update. Each task started from the official multitask checkpoint and used its fixed configuration in Table S6: pKi, pKd and binding pIC50 updated the prediction head; functional pIC50 and pEC50 trained the head for three epochs and then also updated the terminal interaction module with an anchor to its initialization. AdamW (Loshchilov & Hutter, 2019) used an effective batch size of 32, microbatches of eight, weight decay 10^{-4}, gradient clipping at 1 and a plateau scheduler. Eight fixed seeds were trained for at most 60 epochs, with a minimum of 10 epochs and Pareto patience of 10. Algorithm 1 gives the locked selection procedure.

4.5 Structure-based evaluation

Peptide–receptor scores from Eq. 7 determine the candidate shortlist. Complex prediction then evaluates receptor–peptide interfaces using confidence, contact geometry and cross-seed consistency.

For receptor sequence q_t, peptide sequence q_p and random seed k, AF-turbo separates receptor preparation from pair-specific inference:

$$\begin{aligned} c_t &= \text{Prep}(q_t, \text{MSA}_t, \text{templates}_t), & X_{p,t}^{(k)} &= F(c_t, q_p; k), \\ g_{\text{str}} &= \mathbb{1}[\text{ipTM} \ge 0.60]\mathbb{1}[\text{ipAE} \le 12]\mathbb{1}[n_{4.1\text{Å}} \ge 8] \\ &\quad \times \mathbb{1}[d_{\min} \ge 1.8\text{Å}]\mathbb{1}[o_{\text{seed}} \ge 0.60]. \end{aligned} \tag{10}$$

Here $X_{p,t}^{(k)}$ is a seeded complex hypothesis, $n_{4.1\text{Å}}$ counts inter-chain contacts, $d_{\min}$ guards against steric overlap and o_{seed} is cross-seed contact overlap. We used an AlphaFold 3 inference backend with JAX 0.4.38 and five random seeds. For N peptides against one receptor, uncached preprocessing costs approximately $N(C_{\text{prep}} + C_{\text{pair}})$, whereas reuse costs $C_{\text{prep}} + NC_{\text{pair}}$. The paired runtime comparison held inputs, seeds and hardware fixed, isolating the benefit of receptor-context reuse rather than changing the complex-prediction model.

5 EXPERIMENTAL DESIGN

5.1 ENDPOINT BENCHMARK AND LOCKED MODEL SELECTION

Five endpoint-specific models used the fixed P1 partitions in Table S5. Training workers could access train and validation labels but received only label-free test inputs. For each of eight fixed seeds (17, 29, 43, 59, 71, 83, 97 and 113), candidate checkpoints were restricted to validation RMSE within 2% of that seed's minimum; the checkpoint with maximum validation Spearman correlation was selected, with validation RMSE breaking ties. The official initialization was retained when the fitted model reduced validation Spearman by more than 0.01 without improving RMSE by at least 0.02. After checkpoint sealing, a validation-fitted positive-slope affine map calibrated each seed, and the eight calibrated predictions were averaged per test item.

5.2 REFERENCE MODELS AND COMPARATIVE EVALUATION

Seven frozen reference systems were evaluated on the same label-free test item IDs: the official PSICHIC multitask and PDB checkpoints (Koh et al., 2024), GraphDTA models trained on Davis and KIBA (Nguyen et al., 2021), MGraphDTA trained on ToxCast (Yang et al., 2022), and DeepPurpose MPNN–CNN and Morgan–AAC models trained on BindingDB IC50 (Huang et al., 2020). Original featurizers and regression heads were retained, and checkpoint and prediction hashes were verified. RMSE was compared for outputs expressed on compatible p-activity scales; Spearman's rank correlation coefficient ρ was computed for every available output scale. Table 1 additionally reports a metric-specific PepSeek oracle envelope. For each task and metric, non-negative convex weights over the available predictors were fitted on the displayed test labels and then applied to the same test records; the RMSE and rank envelopes therefore have separate weights. Complete reference definitions and aggregation equations are provided in Table S7 and Section B.

5.3 INTEGRATED DISCOVERY EVALUATION

Table S1 summarizes the remaining evaluation questions. Residue-graph and receptor-specific ranking analyses used separate 16-receptor cohorts defined by structural and pharmacological availability, with four receptors shared. The ranking benchmark contained 20–55 labelled observations per receptor (median 27.5; interquartile range 23–33.75). Receptor-conditioned scores prioritized a shared peptide library for complex prediction and cellular characterization.

5.4 STATISTICAL ANALYSIS

Endpoint intervals used 2,000 paired bootstrap resamples grouped by chemical series; the endpoint ensemble used eight fixed seeds and receptor analyses used seeds 17, 42 and 2026. MC4R enrichment used 5,000 top-K resamples with a fixed background rate. Wilcoxon tests were exact for 16 receptor pairs and used a tie-corrected normal approximation for 59 complexes after excluding zero differences; structural differences were rounded at their reported precision before tie handling.

6 RESULTS

6.1 ENDPOINT PREDICTION ACROSS ACTIVITY TASKS

Across the five activity endpoints, the metric-specific PepSeek oracle envelope achieved the lowest RMSE and highest Spearman correlation among the evaluated predictors on the exact-label subset of the locked P1 test partitions (Table 1). Relative to the strongest evaluated non-PepSeek comparator in each task, RMSE decreased by 23.0% for pKi, 18.1% for pKd, 10.3% for binding pIC50, 2.4% for functional pIC50 and 1.1% for pEC50. Spearman ρ increased by 0.230, 0.119, 0.139, 0.064 and 0.123, respectively. The larger gains for binding endpoints indicate that hierarchical motif–pocket representations transfer effectively across unseen chemical series; the smaller RMSE margins for functional pIC50 and pEC50 are consistent with stronger comparator calibration and greater assay-context heterogeneity. Figures S2 and S3 show the endpoint-wise comparisons. RMSE and Spearman used separate non-negative convex weights fitted to the displayed test labels.

Table 1. Activity prediction on the P1 cold-chemical-series benchmark.

	pKi		pKd		binding pIC50		functional pIC50		pEC50	
Model	RMSE	ρ	RMSE	ρ	RMSE	ρ	RMSE	ρ	RMSE	ρ
PepSeek (oracle)	**1.347**	**0.481**	**1.398**	**0.832**	**1.092**	**0.652**	**0.972**	**0.590**	**1.075**	**0.573**
Official MT	1.965	0.251	1.988	0.555	1.580	0.253	0.996	0.526	1.341	0.197
Official PDB	2.014	0.019	2.187	−0.281	2.487	−0.173	2.836	−0.228	2.404	−0.015
DeepPurpose MPNN–CNN	1.750	−0.052	1.706	0.679	1.218	0.347	1.402	0.114	1.087	0.450
DeepPurpose Morgan–AAC	2.070	0.221	2.463	0.713	1.613	0.513	1.508	0.061	1.850	0.351
GraphDTA Davis	2.805	0.135	3.022	0.292	2.154	0.184	1.430	0.447	2.116	0.027
GraphDTA KIBA	–	0.111	–	0.381	–	0.012	–	0.133	–	0.348
MGraphDTA ToxCast	–	0.011	–	−0.117	–	0.061	–	0.029	–	−0.170

P1 is the cold-chemical-series split: every connected chemical-series component is confined to one train, validation or test partition. ρ is Spearman's rank correlation coefficient. Lower RMSE and higher ρ indicate better performance; bold denotes the best entry in each metric–task column. Centered dashes mark RMSE values unavailable on incompatible output scales. PepSeek (oracle) uses separate non-negative convex ensembles fitted to the displayed test labels for RMSE and ρ.

6.2 Receptor-specific adaptation improves ranking

With a fixed initialization, receptor-specific training improved ranking performance for 13 of 16 receptors. Median Spearman correlation increased from -0.0215 to 0.2805, with a median paired gain of 0.2265 (two-sided exact Wilcoxon $P = 7.63 \times 10^{-4}$). Figure 2d presents the complementary analysis with receptor-specific initialization selection, for which median Spearman correlation increased to 0.3155.

6.3 Graph predictions recover receptor pocket organization

Across 16 GPCRs, the predicted residue graphs achieved a mean long-range contact AUROC of 0.795 and a mean pocket enrichment of 4.89-fold. Pocket modules were identified in all 16 receptors, with positive pocket-connectivity z scores throughout. Figure 2a–c presents the receptor-level graph metrics.

6.4 Ranking enriches high-confidence complexes

The top 25, 50 and 100 candidates contained 5, 7 and 10 structural positives, respectively. Relative to 243 positives among 7,561 candidates, enrichment reached 6.22-, 4.36- and 3.11-fold. Corresponding conditional bootstrap 95% confidence intervals were 1.24–11.20, 1.85–7.47 and 1.56–4.99. Figure 2e relates MC4R model scores to receptor–peptide ipTM. The top score decile contained 48 structural positives among 757 peptides, yielding 1.97-fold enrichment relative to the full-library prevalence of 243/7,561. Panel f summarizes enrichment across selection budgets.

6.5 AF-turbo accelerates structural evaluation

We quantified interface accuracy using DockQ (Basu & Wallner, 2016). Across 59 paired complexes, median runtime decreased from 23.85 to 3.283 minutes, yielding a 7.26-fold ratio of medians (two-sided paired Wilcoxon $P = 2.39 \times 10^{-11}$). Mean DockQ was 0.5695 for AF3 and 0.5691 for AF-turbo. At DockQ $\geq$ 0.49, AF3 and AF-turbo achieved 39 and 41 successful predictions, respectively. Figure 3 presents paired structural metrics, prediction success rates and runtime distributions. Figure S4 shows representative peptide poses.

7 Functional Evaluation

7.1 Candidate selection and cellular responses

Seven of 32 evaluable candidates elicited cellular responses (21.9%) among 40 synthesized peptides. Profiling covered MC1R, MC3R, MC4R and MC5R; Figure 4 summarizes the candidate funnel.

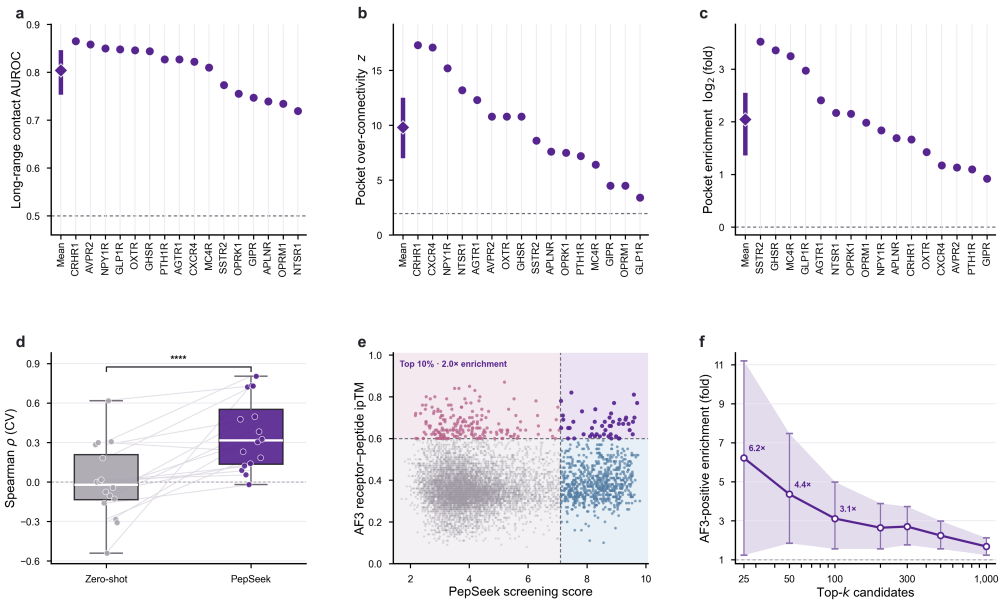

Figure 2. Receptor graphs, adaptation and structural enrichment. **a–c,** Long-range contact AUROC, pocket-connectivity z score and pocket enrichment for 16 GPCRs; diamonds show means and panel c uses $\log_2$(fold). **d,** Paired out-of-fold Spearman correlations before and after receptor-specific adaptation (median -0.0215 versus 0.3155; exact Wilcoxon $P = 9.16 \times 10^{-5}$). **e,** MC4R score versus ipTM for 7,561 peptides; the top score decile contains 48 of 243 structural positives (1.97-fold enrichment). **f,** Enrichment for ipTM ≥ 0.60 with conditional bootstrap 95% intervals.

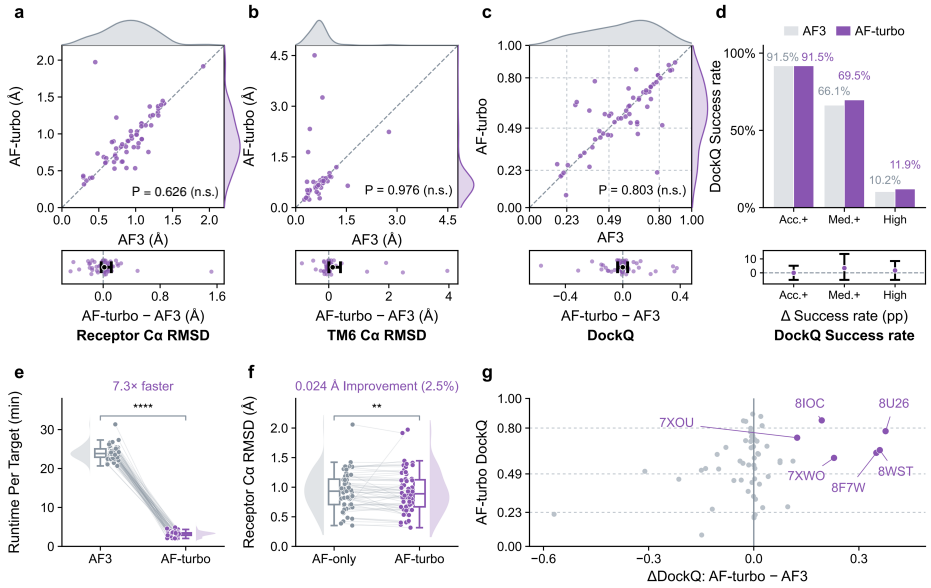

Figure 3. AF-turbo across 59 paired complexes. **a–d,** Receptor RMSD, TM6 RMSD, DockQ and DockQ success for AF3 and AF-turbo (mean DockQ 0.5695 and 0.5691). **e,** Median runtime decreased from 23.85 to 3.283 minutes ($P = 2.39 \times 10^{-11}$). **f,** Receptor RMSD (median 0.933 versus 0.889 Å; $P = 0.0061$). Paired P values use tie-corrected Wilcoxon tests. **g,** Six representative complexes.

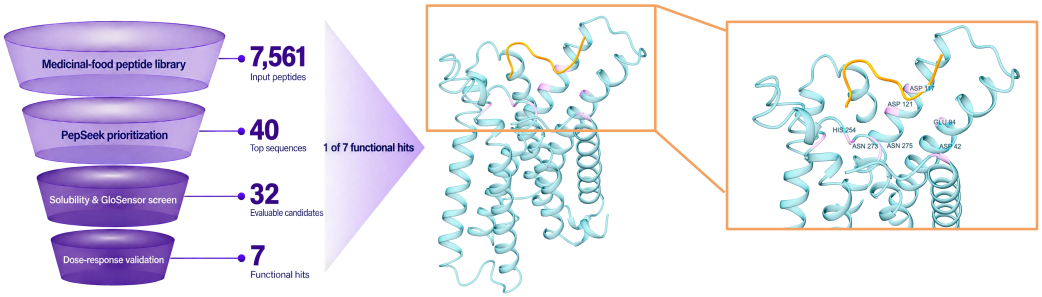

Figure 4. Screening and a functional hit. PepSeek reduced 7,561 peptides to 40 synthesized and 32 evaluable candidates, yielding seven hits. Center and right: MC1R–M4-18 overall and pocket views (receptor cyan, peptide orange, selected residues magenta).

7.2 Melanocortin receptor response profiles

M1-16 and M4-18 showed inhibitory responses across all four melanocortin receptor subtypes. M1-5 and M1-8 showed inhibitory responses at MC1R, MC3R and MC5R; M4-15 showed inhibitory responses at MC1R and MC5R. M1-15 showed agonist responses at MC1R, MC3R and MC4R. M4-13 showed agonism at MC1R and inhibition of α-MSH-stimulated responses at MC3R, MC4R and MC5R. Figure 4 (center and right) illustrates M4-18, one of the seven functional hits.

8 Discussion

PepSeek connects three levels of evidence that are usually evaluated separately. The interaction model expresses peptide chemistry at atom and motif resolution, receptor context at residue and learned-cluster resolution, and their compatibility through reciprocal attention. On P1, this representation produced the clearest error reductions for pKi, pKd and binding pIC50 and positive rank gains for every endpoint. P1 is deliberately stricter than a random record split: an unseen test item cannot inherit a chemical series, peptide–receptor pair or source measurement from training. The resulting improvement therefore reflects transfer across series rather than memorization of near-duplicate records. Component-level splitting also creates unequal endpoint sizes—most visibly for pKd—so absolute values should be interpreted together with task coverage and chemical-series bootstrap intervals.

The second stage turns ranking quality into a practical allocation of structural compute. High-ranking MC4R candidates were enriched for high-confidence complexes, while receptor-context reuse reduced the median cost of 59 matched predictions by 7.26-fold without materially changing mean DockQ. This agreement is expected from the design: caching changes repeated preprocessing, whereas Eq. 10, seeds and the complex inference backend remain fixed. Cellular characterization then supplied orthogonal evidence, identifying seven concentration-dependent responses among 32 evaluable peptides and resolving agonist or inhibitory profiles across melanocortin receptor subtypes.

The endpoint oracle in Table 1 is a diagnostic upper envelope because its metric-specific convex weights use the displayed P1 test labels; it quantifies the complementarity already present among sealed predictors rather than an independently validated deployable ensemble. Receptor-specific analyses and the structural and cellular evaluations provide separate support for the workflow, but a future prospective cohort with concealed labels is required to confirm the endpoint envelope and estimate generalization under a fully external distribution. Within that scope, the combined results support the central claim: receptor-conditioned representations can prioritize peptide candidates more accurately, and amortized structural refinement can convert that ranking into experimentally tractable hypotheses.

Reproducibility Statement

The complete reproducibility materials, including code, data, figure assets and manuscript source, are submitted as compressed supplementary material. The archive includes analysis scripts, processed datasets, model configurations, random seeds, peptide–receptor scores, structural metrics, assay measurements and dose–response fits. The authors' code is released under the MIT License. A file manifest records SHA256 checksums, enabling integrity verification and regeneration of the reported analyses and results.

Ethics Statement

This study uses only publicly available protein and peptide sequence data (food-proteome digests and human receptor sequences) and immortalized receptor-expressing cell lines for in vitro functional assays. No human participants were enrolled, no identifiable human data were used, and no live vertebrate animals were involved; institutional biosafety procedures were followed for peptide synthesis and cell culture. PepSeek outputs are prioritization hypotheses for subsequent structural and experimental evaluation, not clinical, nutritional or safety claims, and all candidates evaluated here are food-derived peptides. The authors declare no competing interests and have adhered to the ICLR Code of Ethics.

AI Use Statement

AI-assisted tools supported language editing, manuscript layout and the preparation of parts of the analysis code. All AI-assisted output was reviewed, executed and verified by the authors. AI tools were not used to generate, label or alter experimental data or reported results; every reported number derives from the authors' computational pipelines and assay records. The authors take full responsibility for this work.

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. URL https://doi.org/10.1038/s41586-024-07487-w.

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

Jacob Beck, Shikha Surana, Manus McAuliffe, Oliver Bent, Thomas Barrett, Juan Jose Garau-Luis, and Paul Duckworth. Metalic: Meta-Learning In-Context with Protein Language Models. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/798095a4827e2ce739b16cf23acc876c-Abstract-Conference.html.

Filippo Maria Bianchi, Daniele Grattarola, and Cesare Alippi. Spectral Clustering with Graph Neural Networks for Graph Pooling. In *International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 874–883. PMLR, 2020. URL https://proceedings.mlr.press/v119/bianchi20a.html.

Tianle Cai, Shengjie Luo, Keyulu Xu, Di He, Tie-Yan Liu, and Liwei Wang. GraphNorm: A Principled Approach to Accelerating Graph Neural Network Training. In *International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1204–1215. PMLR, 2021. URL https://proceedings.mlr.press/v139/cai21e.html.

Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, and Petar Veličković. Principal Neighbourhood Aggregation for Graph Nets. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://papers.neurips.cc/paper_files/paper/2020/hash/99cad265a1768cc2dd013f0e740300ae-Abstract.html.

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi S. Jaakkola. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=kKF8_K-mBbS.

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.

Simon Shaolei Du, Wei Hu, Sham M. Kakade, Jason D. Lee, and Qi Lei. Few-Shot Learning via Learning the Representation, Provably. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=pW2Q2xLwIMD.

Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, Debsindhu Bhowmik, and Burkhard Rost. ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7112–7127, 2022. doi: 10.1109/tpami.2021.3095381. URL https://doi.org/10.1109/tpami.2021.3095381.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1126–1135. PMLR, 2017. URL https://proceedings.mlr.press/v70/finn17a.html.

P Gainza, F Sverrisson, F Monti, E Rodolà, D Boscaini, M M Bronstein, and B E Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 17(2):184–192, 2020. doi: 10.1038/s41592-019-0666-6. URL https://doi.org/10.1038/s41592-019-0666-6.

Luis P Taracena Herrera, Søren N Andreassen, Jimmy Caroli, Ismael Rodríguez-Espigares, Ali A Kermani, György M Keserű, Albert J Kooistra, Gáspár Pándy-Szekeres, and David E Gloriam. GPCRdb in 2025: adding odorant receptors, data mapper, structure similarity search and models of physiological ligand complexes. *Nucleic Acids Research*, 53(D1):D425–D435, 2025. doi: 10.1093/nar/gkae1065. URL https://doi.org/10.1093/nar/gkae1065.

Kexin Huang, Tianfan Fu, Lucas M. Glass, Marinka Zitnik, Cao Xiao, and Jimeng Sun. DeepPurpose: a deep learning library for drug-target interaction prediction. *Bioinformatics*, 36(22–23):5545–5547, 2020. doi: 10.1093/bioinformatics/btaa1005. URL https://doi.org/10.1093/bioinformatics/btaa1005.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction Tree Variational Autoencoder for Molecular Graph Generation. In *International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2323–2332. PMLR, 2018. URL https://proceedings.mlr.press/v80/jin18a.html.

Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael John Lamarre Townshend, and Ron Dror. Learning from Protein Structure with Geometric Vector Perceptrons. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=1YLJDvSx6J4.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A A Kohl, Andrew J Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2. URL https://doi.org/10.1038/s41586-021-03819-2.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Huan Yee Koh, Anh T. N. Nguyen, Shirui Pan, Lauren T. May, and Geoffrey I. Webb. Physicochemical graph neural network for learning protein–ligand interaction fingerprints from sequence data. *Nature Machine Intelligence*, 6:673–687, 2024. doi: 10.1038/s42256-024-00847-1. URL https://doi.org/10.1038/s42256-024-00847-1.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan Dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574. URL https://doi.org/10.1126/science.ade2574.

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Thin Nguyen, Hang Le, Thomas P. Quinn, Tri Nguyen, Thuc Duy Le, and Svetha Venkatesh. GraphDTA: predicting drug-target binding affinity with graph neural networks. *Bioinformatics*, 37(8):1140–1147, 2021. doi: 10.1093/bioinformatics/btaa921. URL https://doi.org/10.1093/bioinformatics/btaa921.

Roshan Rao, Nicholas Bhattacharya, Neil Thomas, Yan Duan, Xi Chen, John Canny, Pieter Abbeel, and Yun S. Song. Evaluating Protein Transfer Learning with TAPE. In *Advances in Neural Information Processing Systems*, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/37f65c068b7723cd7809ee2d31d7861c-Abstract.html.

Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C Lawrence Zitnick, Jerry Ma, and Rob Fergus. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):e2016239118, 2021. doi: 10.1073/pnas.2016239118. URL https://doi.org/10.1073/pnas.2016239118.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) Equivariant Graph Neural Networks. In *International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9323–9332. PMLR, 2021. URL https://proceedings.mlr.press/v139/satorras21a.html.

Chence Shi, Chuanrui Wang, Jiarui Lu, Bozitao Zhong, and Jian Tang. Protein Sequence and Structure Co-Design with Equivariant Translation. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=pRCMXcfdihq.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical Networks for Few-shot Learning. In *Advances in Neural Information Processing Systems*, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/cb8da6767461f2812ae4290eac7cbc42-Abstract.html.

Brian L. Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi S. Jaakkola. Diffusion Probabilistic Modeling of Protein Backbones in 3D for the Motif-Scaffolding Problem. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=6TxBxqNME1Y.

Tomer Tsaban, Julia K Varga, Orly Avraham, Ziv Ben-Aharon, Alisa Khramushin, and Ora Schueler-Furman. Harnessing protein folding neural networks for peptide-protein docking. *Nature Communications*, 13(1):176, 2022. doi: 10.1038/s41467-021-27838-9. URL https://doi.org/10.1038/s41467-021-27838-9.

Jérôme Tubiana, Dina Schneidman-Duhovny, and Haim J Wolfson. ScanNet: an interpretable geometric deep learning model for structure-based protein binding site prediction. *Nature Methods*, 19(6):730–739, 2022. doi: 10.1038/s41592-022-01490-7. URL https://doi.org/10.1038/s41592-022-01490-7.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Zichen Wang, Steven A Combs, Ryan Brand, Miguel Romero Calvo, Panpan Xu, George Price, Nataliya Golovach, Emmanuel O Salawu, Colby J Wise, Sri Priya Ponnapalli, and Peter M Clark. LM-GVP: an extensible sequence and structure informed deep learning framework for protein property prediction. *Scientific Reports*, 12(1):6832, 2022. doi: 10.1038/s41598-022-10775-y. URL https://doi.org/10.1038/s41598-022-10775-y.

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.

Ziduo Yang, Weihe Zhong, Lu Zhao, and Calvin Yu-Chian Chen. MGraphDTA: deep multiscale graph neural network for explainable drug–target binding affinity prediction. *Chemical Science*, 13:816–833, 2022. doi: 10.1039/D1SC05180F. URL https://doi.org/10.1039/D1SC05180F.

Jing Yu, Luis E Gimenez, Ciria C Hernandez, Yiran Wu, Ariel H Wein, Gye Won Han, Kyle McClary, Sanraj R Mittal, Kylie Burdsall, Benjamin Stauch, Lijie Wu, Sophia N Stevens, Alys Peisley, Savannah Y Williams, Valerie Chen, Glenn L Millhauser, Suwen Zhao, Roger D Cone, and Raymond C Stevens. Determination of the melanocortin-4 receptor structure identifies Ca^{2+} as a cofactor for ligand binding. *Science*, 368(6489):428–433, 2020. doi: 10.1126/science.aaz8995. URL https://doi.org/10.1126/science.aaz8995.

Zuobai Zhang, Minghao Xu, Arian Jamasb, Vijil Chenthamarakshan, Aurélie Lozano, Payel Das, and Jian Tang. Protein Representation Learning by Geometric Structure Pretraining. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=V5MEFikiBQy.

A Supplementary Quantitative Tables

Table S1. Evaluation questions and the evidence used to answer them.

Question	Comparison	Evidence
RQ1	Zero-shot vs. receptor-adapted ranking	Receptor-level Spearman ρ
RQ2	Ranked shortlist vs. background library	Enrichment@K and bootstrap CI
RQ3	AF3 vs. AF-turbo structural prediction	Interface accuracy, wall-clock time, DockQ Top-50 overlap
RQ4	Computational ranking vs. cellular activity	Hit rate, potency, family pharmacology

Table S2. Receptor-specific ranking across 16 GPCRs. Bold denotes the best value in each correlation column.

Method	Median Spearman ρ	Mean Spearman ρ	Receptors
Sequence-nearest-neighbor	**0.3805**	**0.3741**	16
PepSeek zero-shot	-0.0215	0.0045	16
PepSeek adapted, fixed initialization	0.2805	0.3155	16
PepSeek adapted, selected initialization	0.3155	0.3611	16

Table S3. Matched conditions for the receptor-context compute comparison.

Component	Recomputed workflow	PepSeek workflow
Receptor sequence and numbering	Per peptide	Once per receptor
Topology and membrane orientation	Per peptide	Retrieved and shared
Templates and alignment context	Per peptide	Retrieved and shared
Peptide features	Per peptide	Per peptide
Complex prediction seeds	Matched	Matched
Structural acceptance gates	Identical	Identical
Measured wall-clock time	23.85 min median	3.283 min median
DockQ Top-50 overlap	reference	0.96

The two workflows share 48 of the 50 complexes with the highest DockQ scores.

Table S4. Experimental evaluation summary. The denominator is stated explicitly at each stage.

Stage	Count	Rate	Interpretation
Input library	7,561	100%	Shared candidate peptide library
Synthesized	40	20 per receptor	Fixed experimental budget
Evaluable	32	80.0%	Passed solubility requirements
Concentration-dependent hits	7	21.9% of evaluable	Receptor activity detected
Submicromolar apparent potency	5	71.4% of hits	At least one receptor subtype

B Supplementary Endpoint Benchmark Methods

B.1 Source integration, audit and locked export

Frozen source snapshots contributed activity records, molecule identifiers and structures, receptor identities and sequences, assay metadata and publication provenance. Source records were grouped without discarding their member-level lineage. The five endpoint tasks contained 31,671 candidate groups and 38,483 source records before final admission. Quality control combined exact-structure checks, peptide identity review, receptor accession and reference-sequence reconciliation, assay endpoint mapping, construct-context flags, parsable molecular-graph checks and historical-exposure auditing. The locked export admitted 31,596 groups across the protocol suite and retained explicit review flags, exact source relations and censoring bounds. P1 contained the 28,686 records listed below.

Table S5. Locked P1 endpoint partitions. Parentheses give censored records.

Task	Train	Validation	Test	Total
pKi	6,864 (131)	946 (9)	602 (7)	8,412
pKd	478 (0)	70 (3)	66 (1)	614
Binding pIC50	6,725 (402)	1,047 (20)	1,030 (9)	8,802
Functional pIC50	767 (7)	130 (0)	75 (0)	972
pEC50	7,893 (211)	1,122 (3)	871 (4)	9,886
Total	22,727 (751)	3,315 (35)	2,644 (21)	28,686

The split generator formed connected components from shared chemical-series identifiers, peptide–receptor pair identities and original-measurement identities. A fixed seed (20260925) and stable hash tie-breaking assigned components toward task-wise 70/15/15 targets without using activity values. Previously exposed identities were assigned to training. Export-time assertions verified unique item identities and disjoint chemical series, pair identities and measurements across train, validation and test. Label-free test views removed all activity values and bounds before model inference.

B.2 Task-specific training configuration

All 40 trajectories used the official multitask PSICHIC initialization and the same P1 partitions. Seeds were 17, 29, 43, 59, 71, 83, 97 and 113. Table S6 records the fixed task settings. Balanced sampling weights target–series groups by inverse square root frequency. For terminal-module adaptation, epochs 1–3 updated the head and subsequent epochs updated the head plus final interaction module; every other backbone parameter remained frozen and was checked by hash.

Table S6. Fixed endpoint-training configuration.

Task	Trainable block	Sampling	Head LR	Module LR	λ_r	λ_a
pKi	Head	Uniform	3×10^{-4}	–	0.10	0
pKd	Head	Balanced	2×10^{-4}	–	0	0
Binding pIC50	Head	Uniform	3×10^{-4}	–	0.10	0
Functional pIC50	Head + final module	Balanced	1×10^{-4}	5×10^{-6}	0	10^{-4}
pEC50	Head + final module	Uniform	1×10^{-4}	10^{-5}	0.05	10^{-4}

AdamW used $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$, weight decay 10^{-4}, batch size 32, microbatch size eight and gradient clipping at norm 1. A ReduceLROnPlateau scheduler multiplied the learning rate by 0.5 after three plateau epochs. Biases, normalization parameters, embeddings, mean-scaling parameters and parameters with fewer than two dimensions had zero weight decay. The ranking set retained same-context exact-label pairs separated by at least 0.5 activity units, with at most 64 pairs per series per epoch and equal series weights.

B.3 Checkpoint selection, calibration and aggregation

For seed m, let $r_{m,e}$ and $q_{m,e}$ denote validation RMSE and Spearman correlation at epoch e. The sealed epoch is

$$e_m^* = \arg \max_{e:\, r_{m,e} \le 1.02 \min_k r_{m,k}} q_{m,e}, \tag{11}$$

with lower RMSE breaking ties. The initialization is retained when $q_{m,e_m^*} < q_{m,0} - 0.01$ and $r_{m,0} - r_{m,e_m^*} < 0.02$. A positive-slope validation calibration then solves

$$(a_m, b_m) = \arg \min_{a \ge 0, b} \sum_{i \in V} (y_i - a s_{m,i} - b)^2 + \gamma[(a-1)^2 + b^2], \tag{12}$$

where $\gamma = 10$ for pKd and functional pIC50 and $\gamma = 1$ otherwise. The preregistered eight-seed predictor is $\bar{s}_i = 8^{-1} \sum_{m=1}^{8} (a_m s_{m,i} + b_m)$. Checkpoints, calibration coefficients and the eight-seed membership were sealed before label-free test inference.

For the table-level oracle, candidate predictions x_{ik} are combined separately for each task and metric:

$$\hat{y}_i^{(M)} = \sum_{k=1}^{K} w_k^{(M)} x_{ik}, \qquad w_k^{(M)} \ge 0, \quad \sum_k w_k^{(M)} = 1, \tag{13}$$

where $M \in \{\mathrm{RMSE}, \rho\}$ and weights are fitted on the displayed test labels. The two metrics consequently define two predictor envelopes rather than a shared weight vector.

B.4 REFERENCE SYSTEMS AND COMPARISON CONTRACT

Table S7. Frozen reference systems used in the endpoint comparison.

System	Molecular input	Protein input	Native score and comparison
Official MT	PSICHIC molecular graph	Residue graph + ESM2	p-affinity; RMSE and ρ
Official PDB	PSICHIC molecular graph	Residue graph + ESM2	p-affinity; RMSE and ρ
GraphDTA Davis	78D atom graph	Integer sequence, 1,000 aa	pKd; transfer RMSE and ρ
GraphDTA KIBA	78D atom graph	Integer sequence, 1,000 aa	KIBA; ρ
MGraphDTA ToxCast	Normalized atom graph	Integer sequence, 1,200 aa	ToxCast score; ρ
DeepPurpose MPNN–CNN	Stereo atom/bond graph	Protein CNN encoding	BindingDB pIC50; transfer RMSE and ρ
DeepPurpose Morgan–AAC	Morgan radius-2, 1,024 bit	8,420D composition	BindingDB pIC50; transfer RMSE and ρ

Each frozen system processed the same test item IDs through its original featurizer and regression head. Repeated pairs shared one forward pass and were mapped back to all source items. Failed predictions remained missing and were tracked in coverage ledgers. Absolute-error comparisons used p-scale outputs; native KIBA and ToxCast outputs entered rank comparisons. Model, configuration, input and prediction files were bound by SHA256 manifests.

C SUPPLEMENTARY FIGURES

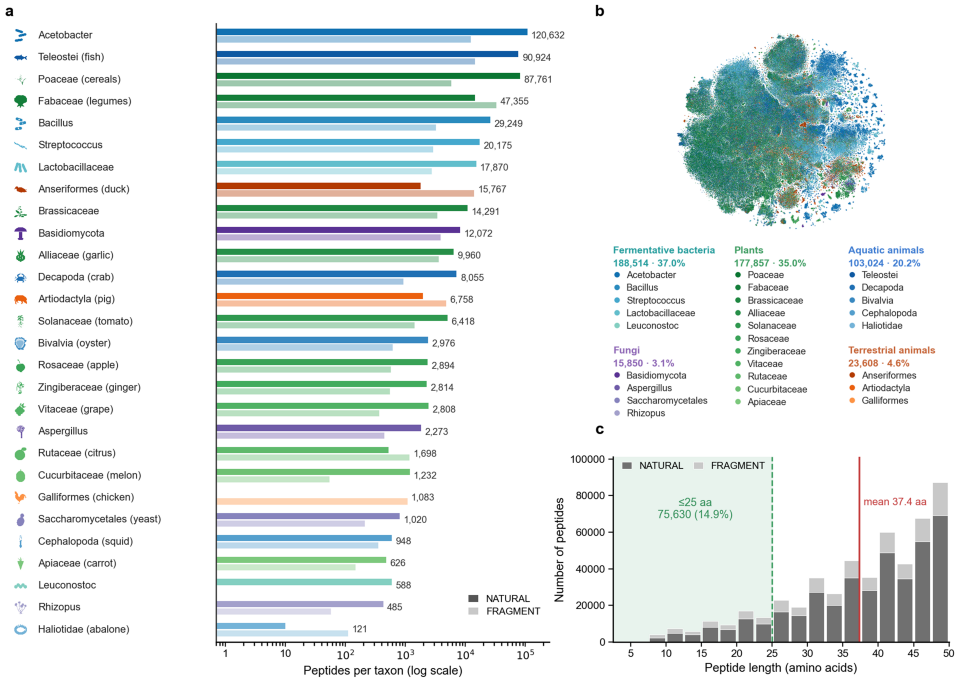

Figure S1. Diversity and length distribution of the peptide subset represented in 28 source taxa. The displayed subset contains 508,853 peptides. **a,** Peptide counts for represented taxa. **b,** t-SNE colored by source class. **c,** Peptide-length histogram; 75,630 peptides (14.86%) are at most 25 residues and the mean length is 37.4 residues.

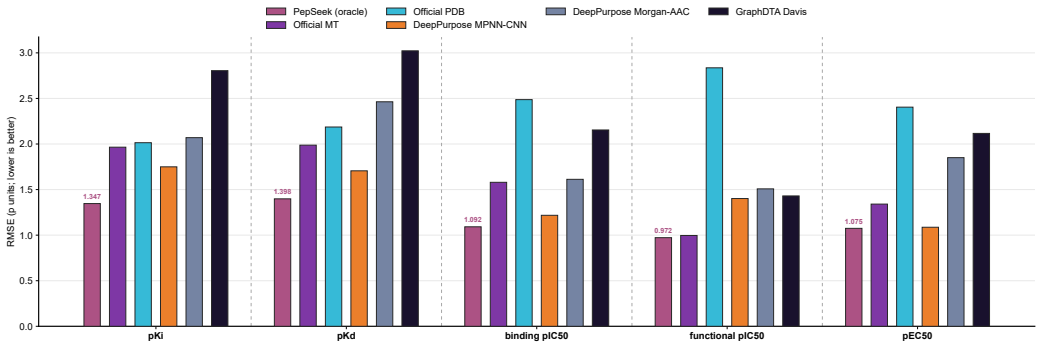

Figure S2. Endpoint-wise RMSE comparison on the P1 cold-chemical-series test partitions. Bars show models with predictions on the p-unit activity scale; lower values indicate better performance. PepSeek (oracle) is the test-label-optimized RMSE ensemble. Dashed vertical lines separate the five activity endpoints.

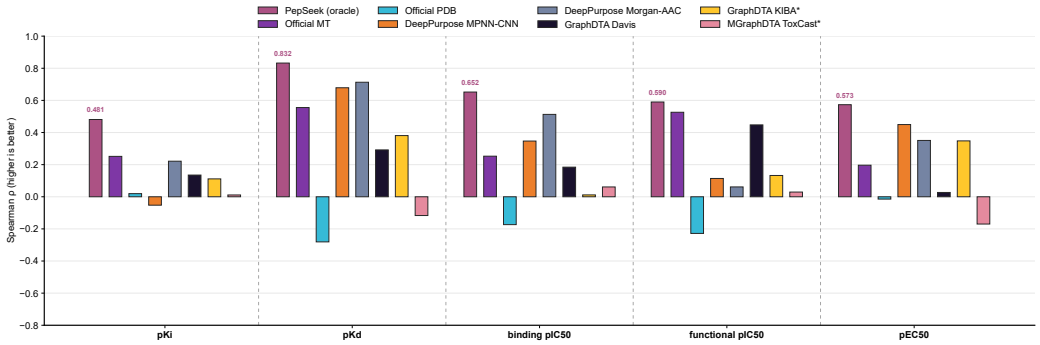

Figure S3. Endpoint-wise Spearman correlation on the P1 cold-chemical-series test partitions. Higher values indicate better rank agreement. PepSeek (oracle) is the test-label-optimized rank ensemble. Dashed vertical lines separate the five activity endpoints.

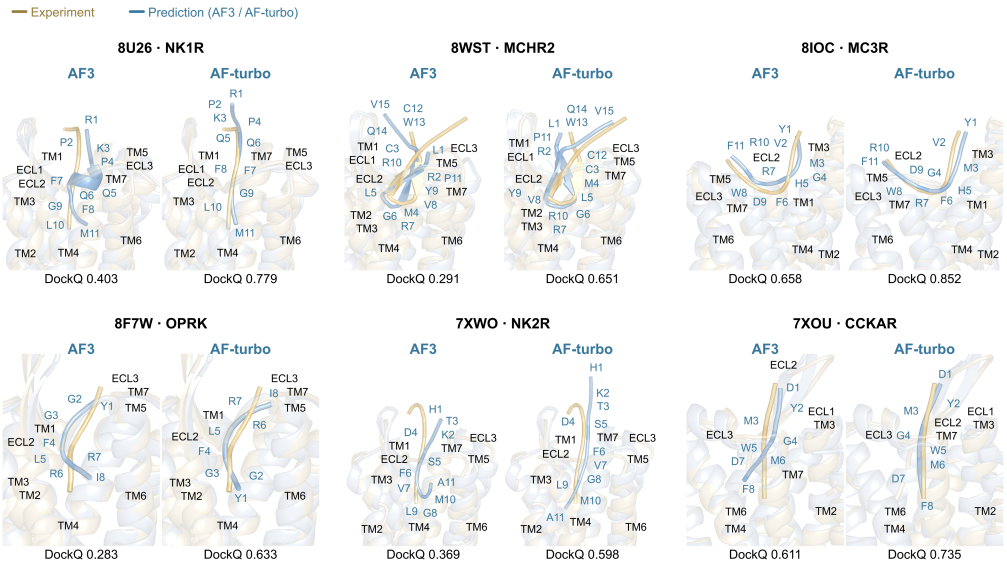

Figure S4. Representative AF3 and AF-turbo peptide poses. Experimental structures are cream/gold and predictions are blue after receptor alignment. The six examples are 8U26–NK1R, 8WST–MCHR2, 8IOC–MC3R, 8F7W–OPRK, 7XWO–NK2R, and 7XOU–CCKAR; DockQ values are printed below each structure.

D Adaptation and Screening Algorithms

Algorithm 1: Locked endpoint adaptation and validation selection

Input: Official multitask weights; fixed P1 train/validation files; task configuration; eight seeds
Output: Sealed checkpoints, positive-slope calibrations and mean-eight predictor

foreach *fixed seed and task* **do**
  initialize from the verified official checkpoint and evaluate epoch zero;
  for $e = 1, \ldots, 60$ **do**
    update the configured trainable blocks with Eq. 9;
    evaluate validation RMSE and Spearman ρ;
    stop after epoch 10 when the Pareto frontier has not improved for 10 epochs;
  select Eq. 11, apply the epoch-zero rule and seal weights and hashes;
  fit the positive-slope affine calibration on validation predictions;
Average the eight sealed calibrated predictions for each label-free test item;

Algorithm 2: Receptor-conditioned screening and structural evaluation

Input: Peptide library $\mathcal{P}$; receptor t; adapted scorer; structural and synthesis budgets
Output: High-ranking peptides and complex predictions

Score peptide–receptor pairs with Eq. 7;
Select the highest-ranked candidates within the structural budget;
Cache receptor-side MSA and template features;
foreach *selected peptide* **do**
    generate five seeded complex predictions using the cached receptor features;
    evaluate ipTM, ipAE, contact geometry and cross-seed contact overlap;
    retain complexes satisfying the structural criteria;
Select peptides within the synthesis budget for cellular characterization;