SITE-AWARE RECEPTOR-SEQUENCE-CONDITIONED PEPTIDE BINDER GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Sequence-based peptide binder generation enables efficient design from readily available receptor sequences. However, binding alone is insufficient for functional peptides, whose biological activity often depends on engaging specific receptor regions. Existing sequence-based methods either offer no direct control over which receptor site is engaged or require the target region to be specified in advance, limiting their use when the functionally relevant region is unknown. To address this gap, we introduce **SitePepGen**, a **Site**-aware, receptor-sequence-conditioned masked language model for **Pep**tide **Gen**eration that requires neither receptor structures nor predefined functional binding sites. SitePepGen combines two components: (1) a residue-level receptor binding-site prediction head, jointly trained with peptide generation using shared ESM-2 representations to identify putative peptide-binding regions; and (2) a site-aware candidate selection module that ranks generated peptides by their physicochemical compatibility score with the predicted receptor sites and the model's generation confidence. In the evaluation of receptor binding-site recovery using predicted receptor–peptide complexes, SitePepGen improves macro-averaged Jaccard and F1 over the strongest baseline by 3.62 and 3.56 percentage points, respectively, while maintaining comparable sequence-plausibility diagnostics. For receptor binding-site prediction, SitePepGen further improves AUROC and AUPRC over the strongest baseline by 0.95 and 1.69 percentage points, respectively. The code will be released at `https://anonymous.4open.science/r/SitePepGen/`.

1 INTRODUCTION

Peptide binders combine high specificity, tunable chemistry, and the ability to engage protein surfaces that are often difficult to target with conventional small molecules (Muttenthaler et al., 2021). These properties make peptides valuable in chemical biology and therapeutic development. We build on PepMLM, a receptor-sequence-conditioned peptide binder generator (Chen et al., 2025a). Unlike approaches that rely on complex structures, this setting does not require an experimentally resolved receptor–peptide complex.

For peptides intended to modulate a specific protein–protein interaction, binding to the receptor alone may not be sufficient to produce the intended effect; engagement of the relevant interaction interface also matters (van Wier & Beekman, 2025). The receptor residues involved in binding a generated peptide are therefore an important design consideration in addition to sequence plausibility and predicted complex confidence. Nevertheless, binding location remains implicit in many sequence-based generation methods.

Existing methods differ in how they account for where a peptide binds. Sequence-based methods such as PepMLM generate peptides from receptor sequences but do not explicitly model receptor binding sites (Chen et al., 2025a). Motif-guided methods such as moPPIt-v3 guide binding toward a predefined target motif but depend on that site information being supplied in advance (Chen et al., 2025b). Structure-guided methods such as EvoBind2 (Li et al., 2025b), PepHAR (Li et al., 2025a), and PepTri (Nguyen et al., 2026) use structure prediction or receptor geometry to guide design. Overall, sequence-based generation typically leaves receptor binding sites implicit, whereas guidance toward a binding region often requires additional motif or structural information. We therefore

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.

investigate how a model can account for receptor-site information during peptide generation without requiring a receptor structure or predefined target site at inference.

We introduce SitePepGen, a site-aware receptor-sequence-conditioned framework for peptide binder generation. SitePepGen augments a receptor-conditioned masked language model with a residue-level receptor-site prediction head on a shared ESM-2 representation. The model learns peptide generation together with receptor binding-site prediction and uses predicted site evidence alongside peptide likelihood to prioritize candidate sequences. Together, these components allow SitePepGen to consider both sequence plausibility and potential receptor-binding regions without requiring a receptor structure or predefined target site at inference.

To evaluate SitePepGen's effectiveness in recovering receptor-binding sites, we use ColabFold to predict receptor–peptide complex structures after generation and compare the receptor residues involved in peptide binding in the predicted and reference complexes (Evans et al., 2021; Mirdita et al., 2022). SitePepGen improves macro-averaged Jaccard and F1 for receptor binding-site recovery over PepMLM by 3.62 and 3.56 percentage points, respectively, while maintaining comparable sequence-quality diagnostics.

Our contributions are summarized as follows.

- We introduce SitePepGen, to our knowledge the first receptor-sequence-conditioned peptide generation framework to jointly learn residue-level receptor binding-site prediction and peptide generation through a shared ESM-2 backbone. By inferring potential receptor binding sites and using them to guide candidate selection, SitePepGen enables site-aware peptide design without requiring a receptor structure or a predefined target site at inference.
- We introduce a site-aware candidate selection module that uses a heuristic physicochemical compatibility score between the predicted receptor binding site and generated peptides to prioritize candidates. It combines this compatibility score with language-model scores without requiring complex-structure prediction during selection.
- We evaluate SitePepGen at two complementary levels. For receptor binding-site recovery from predicted complexes, it improves macro-averaged Jaccard and F1 over PepMLM by 3.62 and 3.56 percentage points, respectively, while maintaining comparable sequence-plausibility diagnostics. In a separate residue-level site-prediction evaluation, its site head improves AUROC and AUPRC over ALLSites by 0.95 and 1.69 percentage points, respectively.

2 Related Work

Receptor-sequence-input peptide design. Protein language models provide sequence priors for peptide generation without requiring an experimentally resolved target structure. PepMLM fine-tunes ESM-2 with span masking to reconstruct a peptide appended to a receptor sequence (Lin et al., 2023; Chen et al., 2025a). PepPrCLIP generates naturalistic peptide candidates by perturbing peptide representations in the ESM-2 latent space and prioritizes them using a contrastive model of target specificity (Bhat et al., 2025). BOND-PEP is a sequence-first, retrieval-augmented generator with topology-conditioned, explicit residue-resolved conditioning, whereas PepCCD is contrastive-conditioned diffusion driven by target-sequence representations (Ding, 2026; Zhang et al., 2026). EvoBind2 also begins from a target sequence alone, but uses structure-prediction-guided directed evolution to search peptide sequences and binding configurations without a predefined binding site (Li et al., 2025b). Together, these methods use method-specific sequence, retrieval/topology-conditioning, diffusion, or structure-prediction mechanisms rather than jointly learning peptide generation with residue-level receptor binding-site prediction.

Target-motif-guided generation provides more direct control over binding location. moPPIt-v3 takes a target sequence together with user-specified target motifs and uses a residue-level interaction predictor and an affinity objective to guide peptide generation (Chen et al., 2025b). This formulation supports motif-specific design when the relevant receptor residues are known, but requires the target region to be specified before inference.

Structure-conditioned peptide design. Structure-conditioned approaches use receptor structures, binding pockets, or hotspot information to impose geometric constraints during peptide design.

Representative methods include PepGLAD, PepFlow, and DiffPepBuilder (Kong et al., 2024; Li et al., 2024; Wang et al., 2024); PepHAR and PepTri provide additional structure-guided peptide-design strategies (Li et al., 2025a; Nguyen et al., 2026). Broader frameworks such as RFdiffusion followed by ProteinMPNN extend the same paradigm to binder-backbone and sequence design (Watson et al., 2023; Dauparas et al., 2022). These approaches provide explicit geometric control but require structural or spatial information at inference.

Receptor binding-site prediction and structural evaluation. Receptor binding-site predictors estimate which receptor residues are likely to interact with a peptide. PepNN predicts peptide-binding residues using learned sequence and structural representations (Abdin et al., 2022). PepCA uses ESM-2 sequence representations and peptide–protein co-attention for residue-level binding-site prediction (Huang et al., 2024). ALLSites extends binding-site prediction across drug modalities, while GeoPep uses a geometry-aware masked language model for protein–peptide site prediction (Mou et al., 2026; Chen et al., 2026). SaLT&PepPr predicts protein-interaction interfaces from sequence and prioritizes peptide motifs derived from natural interaction partners (Brixi et al., 2023). These approaches provide interface information or use it to prioritize peptide motifs, but do not jointly train de novo peptide generation with residue-level receptor-site prediction.

Complex-structure prediction provides a complementary means of evaluating generated binders. AlphaFold-Multimer and ColabFold can predict receptor–peptide complex structures, from which receptor binding-site information can be extracted (Evans et al., 2021; Mirdita et al., 2022). They are structure-prediction tools rather than dedicated interface predictors and can be used either inside a design loop, as in EvoBind2, or after generation for structural assessment.

Taken together, prior work covers blind design from receptor-sequence input, explicit local control through user-specified target motifs, structure-conditioned peptide generation, and standalone receptor binding-site prediction. These approaches make different assumptions about the information and computation available during inference. A remaining question is how to incorporate learned residue-level receptor-site information into peptide generation and candidate selection without requiring a receptor structure or predefined target site at inference, so as to improve recovery of reference receptor binding-site residues by generated peptides. SitePepGen incorporates residue-level receptor-site learning into a sequence generator and uses the learned site evidence during candidate selection. It requires neither a receptor structure nor a predefined target site during generation and selection, while reserving complex-structure prediction for post-generation site-fidelity evaluation.

3 METHOD

3.1 PRELIMINARIES

Let (r, p) denote a receptor–peptide training pair, where $r = (r_1, \ldots, r_j, \ldots, r_M)$ and $p = (p_1, \ldots, p_i, \ldots, p_L)$ are the receptor and paired peptide sequences of lengths M and L, respectively. The receptor binding-site labels are denoted by $z = (z_1, \ldots, z_M) \in \{0, 1\}^M$, where $z_j = 1$ indicates that receptor residue r_j is part of the receptor binding site in the reference complex. For receptor-conditioned generation, L masked peptide positions are appended to r, forming $x = (r, [\mathrm{MASK}]^L)$.

Let $\mathbf{h}_j$ denote the shared sequence representation of receptor residue r_j. The peptide-generation head predicts a position-specific amino-acid distribution $P_\theta(\cdot \mid x)$ at each masked peptide position, while the site-prediction head predicts q_j, the probability that r_j belongs to the receptor binding site. Here θ denotes the model parameters.

3.2 OVERVIEW

Figure 1 provides an overview of SitePepGen. A shared ESM-2 backbone (Lin et al., 2023) encodes the masked input x, and its representations feed a peptide-generation head and a site-prediction head.

The generation head predicts an amino-acid distribution $P_\theta(\cdot \mid x)$ at each masked peptide position. It reconstructs the paired peptide p over the L masked positions, giving the objective

$$\mathcal{L}_{\mathrm{MLM}} = -\frac{1}{L} \sum_{i=1}^{L} \log P_\theta(p_i \mid x). \tag{1}$$

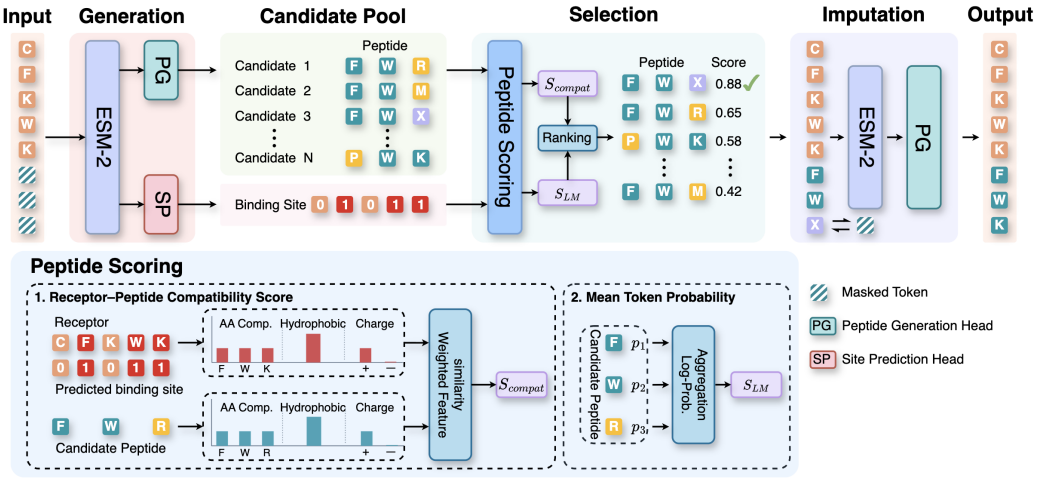


Figure 1: Overview of SitePepGen. Shared ESM-2 representations support receptor binding-site prediction and peptide generation, while predicted site information guides candidate ranking and final peptide selection.

The site-prediction head applies the same linear projection to each receptor-position representation, followed by a sigmoid: $q_j = \sigma(\mathbf{w}_{\mathrm{SP}}^{\top}\mathbf{h}_j + b_{\mathrm{SP}})$, where $\mathbf{w}_{\mathrm{SP}}$ and b_{SP} are learned head parameters. Only receptor-position outputs are used, and q_j estimates the probability that r_j belongs to the receptor binding site. Because the peptide positions in x remain masked, the site probabilities do not depend on a generated candidate peptide. The binding-site labels z_j supervise these predictions across all M receptor residues. To account for class imbalance, the site objective uses a positive-class weight α:

$$\mathcal{L}_{\mathrm{site}} = -\frac{1}{M}\sum_{j=1}^{M}\left[\alpha z_j \log q_j + (1 - z_j)\log(1 - q_j)\right]. \tag{2}$$

The model combines the generation and site objectives as a weighted sum:

$$\mathcal{L} = \lambda_{\mathrm{MLM}}\mathcal{L}_{\mathrm{MLM}} + \lambda_{\mathrm{site}}\mathcal{L}_{\mathrm{site}}. \tag{3}$$

The site objective shapes the shared ESM-2 representations used by the peptide-generation head. The site probabilities are also used to derive a predicted receptor-site mask and normalized site weights, which are shared across the candidate pool.

The generation head constructs a pool of peptide candidates. As detailed in the lower part of Figure 1, the selection module combines site–peptide compatibility with language-model scores to rank the candidates. The imputation step then replaces any unresolved X positions in the selected sequence with canonical amino acids before the final peptide is returned. Candidate generation, selection, and imputation require neither a receptor structure nor a predefined target site.

3.3 Candidate Pool and Site-Aware Candidate Selection

Let $\mathcal{C}$ denote the set of distinct candidate peptides, and let $c = (c_1, \ldots, c_L) \in \mathcal{C}$ denote one candidate. The candidate pool $\mathcal{C}$ retains the initial candidate as a fallback and adds proposals sampled from the frozen generation head, keeping only one copy of each sequence. For each receptor query, the site-prediction head is evaluated once with all peptide positions masked. Its output defines a receptor-site mask and normalized site weights shared by every candidate in $\mathcal{C}$; the site is not recomputed for each peptide.

The receptor–peptide compatibility score compares each candidate with the predicted site in terms of amino-acid composition, hydrophobicity, and charge. For an amino acid a in the set $\mathcal{A}_{20}$ of 20 canonical amino acids, let $f_s(a)$ denote its weighted composition fraction at the predicted receptor site and $f_p(a; c)$ its composition fraction in candidate c. The subscripts s and p stand for *site* and

peptide. Likewise, h_s and $h_p(c)$ denote hydrophobic-residue fractions, while Δ_s and $\Delta_p(c)$ denote positively charged fractions minus negatively charged fractions. The compatibility score is

$$S_{\mathrm{compat}}(c) = \sum_{a \in \mathcal{A}_{20}} f_s(a) f_p(a; c) + \gamma \big[h_s h_p(c) - \Delta_s \Delta_p(c) \big]. \tag{4}$$

The first term measures composition agreement; the second rewards hydrophobic agreement and opposite-charge pairing while penalizing like-charge pairing. This is a fixed selection heuristic, not a physical binding-energy estimate. The peptide composition profile is normalized over canonical residues, whereas its hydrophobic and charge fractions use the requested length as their denominator. The precise residue groups and the value of γ are given in Appendix A.2.

The language-model score is the mean log-probability assigned to candidate c by the frozen generation head with all peptide positions masked:

$$S_{\mathrm{LM}}(c) = \frac{1}{L} \sum_{i=1}^{L} \log P_\theta(c_i \mid x). \tag{5}$$

Both scores are standardized separately within $\mathcal{C}$. Candidates are then ranked by

$$U(c) = Z_{\mathrm{compat}}(c) + Z_{\mathrm{LM}}(c), \tag{6}$$

where Z_{compat} and Z_{LM} denote the standardized compatibility and language-model scores. The highest-ranked candidate is selected only if it satisfies the fixed pseudo-perplexity and composition safeguards relative to the retained baseline; otherwise, the baseline is used. Sampling settings, the site threshold, and safeguard thresholds are specified in Appendix A.2.

4 Experiments

4.1 Datasets

Our dataset is based on the receptor–peptide sequence pairs released with PepMLM, and we retain its original training/test split (Chen et al., 2025a). We match these pairs to receptor and peptide PDB chains using PepNN and Propedia records together with the corresponding structures in the Protein Data Bank (Abdin et al., 2022; Martins et al., 2021; Berman et al., 2000). We then extract receptor binding-site labels from the matched reference complexes using the 5 Å heavy-atom criterion.

After retaining records with usable structural matches, we set aside 500 eligible training pairs for validation. The resulting dataset contains 9,327 training pairs, 500 validation pairs, and 202 held-out test queries. Thus, the sequence pairs and initial train/test assignment come from PepMLM, whereas the PDB-chain matching, receptor binding-site labels, and validation split are constructed for this study.

4.2 Baselines

For peptide generation, we compare SitePepGen with PepMLM (Chen et al., 2025a), BOND-PEP (Ding, 2026), PepPrCLIP (Bhat et al., 2025), and PepCCD (Zhang et al., 2026). PepMLM is the original sequence-based model on which our method builds; the other three provide additional peptide-generation comparisons. All generated peptides are evaluated using the protocol described in Section 4.3.

For residue-level receptor binding-site prediction, we compare with ALLSites (Mou et al., 2026), GeoPep (Chen et al., 2026), and PepCA (Huang et al., 2024). ALLSites uses the receptor sequence alone. GeoPep and PepCA receive the true peptide sequence, whereas SitePepGen does not; we therefore treat them as *partner-aware references* rather than input-matched baselines. We use an adapted GeoPep model with ESM-3 base weights, a newly initialized KAN head, and a 5 Å receptor-classification loss, without its original distance loss. PepCA is initialized from the official PepNN-supervised weights and fine-tuned on our training set.

4.3 Evaluation protocol and metrics

For each generated peptide, LocalColabFold predicts a corresponding receptor-peptide complex using AlphaFold2-Multimer-v3 under a fixed configuration with a public MSA, one model, and three recycles (Evans et al., 2021; Mirdita et al., 2022). A receptor residue is considered to contact the peptide when any of its heavy atoms lies within 5 Å of a peptide heavy atom. The same contact definition is applied to the corresponding reference complex.

Receptor binding-site recovery is evaluated using macro-averaged Jaccard, precision, recall, and F1. Pseudo-perplexity and amino-acid composition are used as sequence diagnostics, while ipTM summarizes predicted complex confidence. Detailed calculations of the receptor binding-site recovery metrics and sequence and structural diagnostics are provided in Appendix A.6.

4.4 Evaluation of peptide generation

Table 1: Macro receptor binding-site recovery on the held-out test set.

Method	Jaccard ↑	Precision ↑	Recall ↑	F1 ↑
BOND-PEP (Ding, 2026)	0.4445	0.5751	0.5767	0.5556
PepPrCLIP (Bhat et al., 2025)	0.4323	0.5408	0.5764	0.5406
PepCCD (Zhang et al., 2026)	0.4043	0.5355	0.5436	0.5200
PepMLM (Chen et al., 2025a)	0.4666	0.5949	0.6001	0.5770
SitePepGen	**0.5028**	**0.6183**	**0.6458**	**0.6126**

Note: *indicates the best performance (bold), and* *indicates the second-best performance.*

SitePepGen achieves the highest scores on all four macro-averaged metrics. Relative to PepMLM, it improves Jaccard by 3.62 percentage points (query-level paired bootstrap 95% confidence interval (CI): [1.91, 5.43]) and F1 by 3.56 percentage points (95% CI: [1.70, 5.53]). BOND-PEP is the strongest of the other evaluated generation methods; SitePepGen exceeds it by 5.83 points in Jaccard and 5.70 points in F1. All generated sequences are assessed with the same 5 Å structural protocol.

To examine how these gains vary across receptor subsets, we divide the test queries into seen and unseen groups according to whether the complete receptor sequence exactly matches one in the training split, yielding 74 and 128 queries, respectively. Tables 2 and 3 report receptor binding-site recovery for the two groups. Evaluation and macro-averaging follow the overall comparison, with each method summarized over its structurally evaluable queries.

Table 2: Macro receptor binding-site recovery for seen receptors.

Method	Jaccard ↑	Precision ↑	Recall ↑	F1 ↑
BOND-PEP (Ding, 2026)	0.4446	0.5788	0.5739	0.5548
PepPrCLIP (Bhat et al., 2025)	0.4366	0.5493	0.5892	0.5458
PepCCD (Zhang et al., 2026)	0.3981	0.5484	0.5280	0.5142
PepMLM (Chen et al., 2025a)	0.4755	0.6087	0.6030	0.5861
SitePepGen	**0.5095**	**0.6245**	**0.6507**	**0.6158**

Note: *indicates the best performance (bold), and* *indicates the second-best performance.*

Table 3: Macro receptor binding-site recovery for unseen receptors.

Method	Jaccard ↑	Precision ↑	Recall ↑	F1 ↑
BOND-PEP (Ding, 2026)	0.4444	0.5729	0.5783	0.5560
PepPrCLIP (Bhat et al., 2025)	0.4299	0.5359	0.5690	0.5376
PepCCD (Zhang et al., 2026)	0.4078	0.5280	0.5526	0.5233
PepMLM (Chen et al., 2025a)	0.4615	0.5869	0.5984	0.5717
SitePepGen	**0.4990**	**0.6146**	**0.6430**	**0.6108**

Note: *indicates the best performance (bold), and* *indicates the second-best performance.*

SitePepGen achieves the highest values on all four metrics in both groups. Relative to PepMLM, its Jaccard and F1 scores increase by 3.40 and 2.97 percentage points for seen receptors, and by 3.75 and 3.91 points for unseen receptors. Thus, the overall gain in site recovery is not confined to receptors whose sequences exactly match those in the training split.

Together, these results indicate that, compared with the baselines, peptides generated by SitePepGen yield predicted complexes whose receptor contact residues show greater overlap with those in the reference complexes. The gains persist for receptors whose full sequences do not exactly match those in the training set. This improvement in contact-residue overlap is achieved using only the receptor sequence as target-specific input for generation and selection, without requiring a receptor structure or a predefined target site.

4.5 Ablation Study

We isolate the contributions of receptor-site supervision, site-aware selection, and `X` imputation. M0 is the original PepMLM baseline. M1 adds the site-prediction objective; M2 adds Selection to M1, while M3 instead adds Imputation. M4 combines all three components and is the complete SitePepGen method. The compared SitePepGen routes share a candidate pool within each run. Here, **Selection** denotes the site-aware candidate ranking described in Section 3.3, and **Imputation** denotes the fixed post-selection procedure that replaces unresolved `X` positions with canonical amino acids.

Table 4: Component ablation for macro receptor binding-site recovery.

Method	SP	Selection	Imputation	Jaccard ↑	Precision ↑	Recall ↑	F1 ↑
M0	–	–	–	0.4666	0.5949	0.6001	0.5770
M1	✓	–	–	0.4790	0.6061	0.6076	0.5875
M2	✓	✓	–	0.4853	0.6141	0.6166	0.5948
M3	✓	–	✓	0.4942	0.6071	0.6314	0.6018
M4 (SitePepGen)	✓	✓	✓	**0.5028**	**0.6183**	**0.6458**	**0.6126**

***Note**: Values are means across runs. indicates the best performance (bold), and indicates the second-best performance.*

M0 corresponds to PepMLM (Chen et al., 2025a). The multi-run mean for M1 exceeds M0 by 1.24 Jaccard points and 1.05 F1 points. Relative to M1, Selection increases F1 by 0.73 points, while Imputation increases it by 1.43 points. Combining both components yields the best macro scores across all four metrics. M4 exceeds M2 and M3 in F1 by 1.78 and 1.08 points, respectively.

4.6 Receptor Binding-Site Prediction

We additionally evaluate residue-level receptor binding-site prediction on the held-out test split. All methods use the same 5 Å receptor–peptide heavy-atom contact labels; checkpoints and classification thresholds are selected using only the validation split. SitePepGen values aggregate repeated training runs without displaying standard deviations, while the other entries are point estimates.

Table 5: Residue-level receptor binding-site prediction on the held-out test set.

Method	AUROC ↑	AUPRC ↑	F1 ↑	MCC ↑
GeoPep (Chen et al., 2026)	0.9552	0.7719	0.6967	0.6781
PepCA (Huang et al., 2024)	0.9344	0.6858	0.6345	0.5994
ALLSites (Mou et al., 2026)	0.9637	0.8261	0.7697	0.7472
SitePepGen	**0.9732**	**0.8430**	**0.7823**	**0.7610**

***Note**: indicates the best performance (bold), and indicates the second-best performance.*

SitePepGen achieves the highest value among the evaluated methods on all four metrics. Relative to ALLSites, its AUROC, AUPRC, F1, and MCC are higher by 0.95, 1.69, 1.26, and 1.38 percentage points, respectively.

Together, these results indicate that SitePepGen's jointly trained site head provides strong residue-level discrimination using only the receptor sequence as target-specific input, outperforming the input-matched ALLSites baseline. Unlike the peptide-aware reference methods, it does not require the paired peptide sequence to predict putative receptor peptide-binding-site residues. The resulting site probabilities also guide candidate selection, connecting residue-level binding-site prediction to receptor-conditioned peptide generation without requiring a receptor structure or a predefined target region at inference.

4.7 Visualization of SitePepGen

Figure 2 presents ten paired comparisons between PepMLM and SitePepGen, with both methods evaluated against the reference binding site on the same receptor chain in each comparison. In cases such as 2XPO:A, 2BDY:A, and 3GIS:D, SitePepGen recovers a larger portion of the reference site (green) and leaves fewer reference contacts unrecovered (yellow). The magnitude of improvement varies across cases, and additional predicted contacts (red) remain in several examples. These visualizations complement the macro-averaged gains in Table 1 by showing how improved binding-site recovery appears spatially in selected predicted complexes. They also show that recovering more reference contacts does not always eliminate contacts outside the reference site.

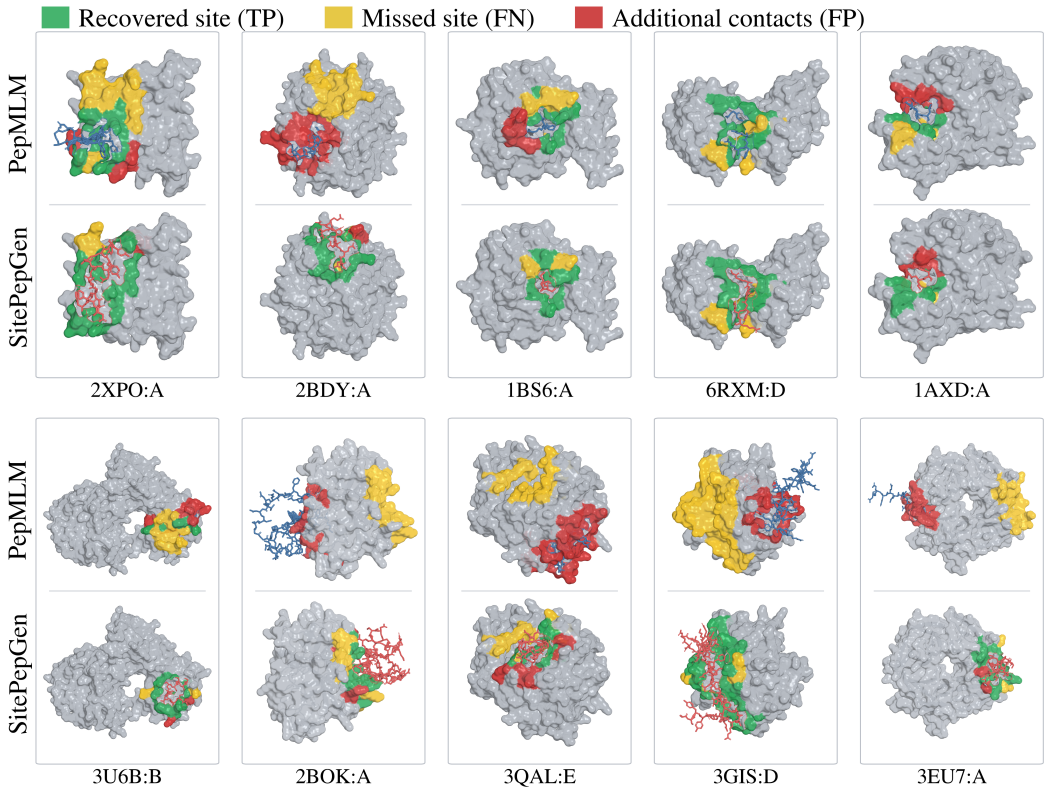


Figure 2: Qualitative receptor binding-site recovery for ten paired test queries under the 5 Å heavy-atom contact definition. The cases are arranged in two groups of five. Within each case, PepMLM is shown above SitePepGen for the same receptor chain. Green denotes recovered reference contacts, yellow denotes missed reference contacts, red denotes additional predicted contacts, and gray denotes the remaining receptor surface.

4.8 Sequence and structural diagnostics

Figure 3 provides complementary analyses of sequence plausibility, receptor conditioning, and predicted complex confidence. In Figure 3(a), the fraction of generated complexes with ipTM at

or above that of the corresponding reference complex is 42.08% for SitePepGen and 38.61% for PepMLM. Although ipTM does not identify the contacted receptor residues, the comparison provides an additional measure of predicted complex confidence alongside the contact-recovery metrics.

Figure 3(b) compares pseudo-perplexity under the matched receptor and a receptor-permuted control. The separation between the two distributions for both generators indicates that the scoring model remains sensitive to receptor context. Figures 3(c–e) show that SitePepGen shifts the pseudo-perplexity distribution toward lower values while maintaining amino-acid composition close to the test distribution. Together, these diagnostics indicate that the improvement in receptor-site recovery is not accompanied by an apparent degradation in sequence plausibility.

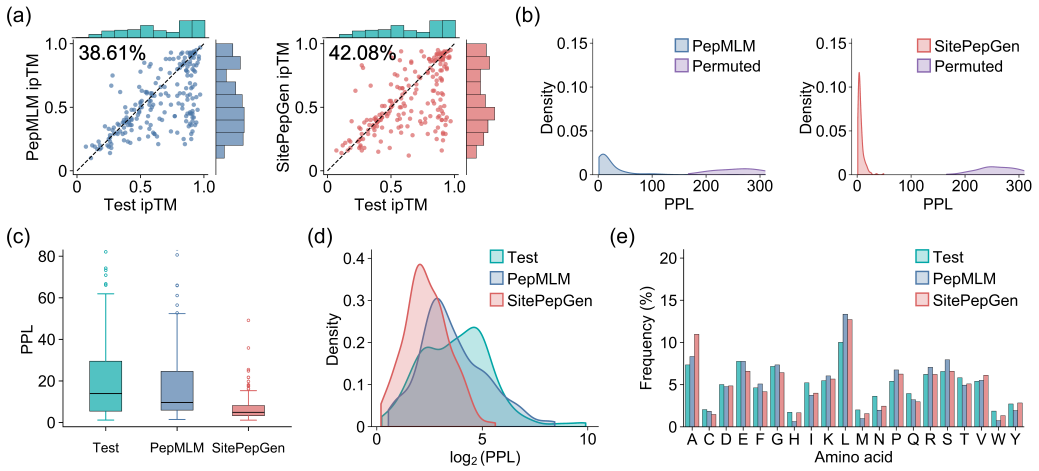


Figure 3: Sequence and structural diagnostics. (a) Generated-complex ipTM compared with the corresponding reference-complex ipTM for PepMLM and SitePepGen. (b) Pseudo-perplexity under matched and receptor-permuted conditions. (c) Pseudo-perplexity boxplots. (d) $\log_2(\mathrm{PPL})$ distributions. (e) Amino-acid composition. Both methods are evaluated using the same frozen scoring model.

5 Conclusion

We introduced SitePepGen, a site-aware peptide generator that jointly learns receptor binding-site prediction and peptide generation and uses the predicted sites to rank candidates. At inference, it requires only the receptor sequence as target-specific input, rather than a receptor structure or a predefined binding region. On 202 held-out queries, SitePepGen improves macro-averaged Jaccard and F1 for receptor binding-site recovery from predicted complexes over PepMLM by 3.62 and 3.56 percentage points, respectively, without apparent loss of sequence plausibility. The gain persists for receptors with no exact sequence match in training. In a separate evaluation, its site-prediction head improves AUROC and AUPRC over the input-matched ALLSites baseline by 0.95 and 1.69 percentage points. Together with the ablation results, these findings support learned receptor-site information as a useful signal for sequence-conditioned peptide design.

AI Use Statement

We used large language models (LLMs) solely for polishing the writing, such as improving sentence fluency and fixing grammar issues. No content generation, idea creation, or experimental analysis was conducted by LLMs.

Ethics Statement

This study uses public protein–peptide records and computationally generated sequences; it involves no human participants, personal data, animal experiments, or clinical intervention. Designed peptides may have unforeseen biochemical activity and should not be synthesized or tested without appropriate biosafety procedures, license review, and target-specific risk assessment. Receptor binding-site overlap, ipTM, and sequence diagnostics are computational proxies and do not establish experimental affinity, safety, or efficacy.

Reproducibility Statement

Section 3 specifies training, candidate generation, scoring, and imputation; Section 4.3 specifies structural evaluation and metrics, while Section 4 defines the comparisons and ablation study. The appendix provides two-stage training settings, dataset statistics, the fixed inference protocol, computational profile, and algorithm. The anonymous repository at `https://anonymous.4open.science/r/SitePepGen/` will provide code and receptor–peptide sequence pairs with binary receptor-site labels.

References

Osama Abdin, Satra Nim, Han Wen, and Philip M. Kim. Pepnn: a deep attention model for the identification of peptide binding sites. *Communications Biology*, 5(1):503, 2022. doi: 10.1038/s42003-022-03445-2. URL https://doi.org/10.1038/s42003-022-03445-2.

Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235. URL https://doi.org/10.1093/nar/28.1.235.

Suhaas Bhat, Kalyan Palepu, Lauren Hong, Joey Mao, Tianzheng Ye, Rema Iyer, Lin Zhao, Tianlai Chen, Sophia Vincoff, Rio Watson, Tian Z. Wang, Divya Srijay, Venkata Srikar Kavirayuni, Kseniia Kholina, Shrey Goel, Pranay Vure, Aniruddha J. Deshpande, Scott H. Soderling, Matthew P. DeLisa, and Pranam Chatterjee. De novo design of peptide binders to conformationally diverse targets with contrastive language modeling. *Science Advances*, 11(4):eadr8638, 2025. doi: 10.1126/sciadv.adr8638. URL https://doi.org/10.1126/sciadv.adr8638.

Garyk Brixi, Tianzheng Ye, Lauren Hong, Tian Wang, Connor Monticello, Natalia Lopez-Barbosa, Sophia Vincoff, Vivian Yudistyra, Lin Zhao, Elena Haarer, Tianlai Chen, Sarah Pertsemlidis, Kalyan Palepu, Suhaas Bhat, Jayani Christopher, Xinning Li, Tong Liu, Sue Zhang, Lillian Petersen, Matthew P. DeLisa, and Pranam Chatterjee. SaLT&PepPr is an interface-predicting language model for designing peptide-guided protein degraders. *Communications Biology*, 6 (1):1081, 2023. doi: 10.1038/s42003-023-05464-z. URL https://doi.org/10.1038/s42003-023-05464-z.

Dian Chen, Yunkai Chen, Tong Lin, Sijie Chen, Levent Burak Kara, and Xiaolin Cheng. GeoPep: A geometry-aware masked language model for protein–peptide binding site prediction. *Journal of Chemical Information and Modeling*, 66(13):7377–7389, 2026. doi: 10.1021/acs.jcim.6c00187. URL https://doi.org/10.1021/acs.jcim.6c00187.

Leo Tianlai Chen, Zachary Quinn, Madeleine Dumas, Christina Peng, Lauren Hong, Moises Lopez-Gonzalez, Alexander Mestre, Rio Watson, Sophia Vincoff, Lin Zhao, Jianli Wu, Audrey Stavrand, Mayumi Schaepers-Cheu, Tian Zi Wang, Divya Srijay, Connor Monticello, Pranay Vure, Rishab Pulugurta, Sarah Pertsemlidis, Kseniia Kholina, Shrey Goel, Matthew P. DeLisa, Jen-Tsan Ashley Chi, Ray Truant, Hector C. Aguilar, and Pranam Chatterjee. Target sequence-conditioned design of peptide binders using masked language modeling. *Nature Biotechnology*, 44(6):1002–1010, 2025a. doi: 10.1038/s41587-025-02761-2. URL https://doi.org/10.1038/s41587-025-02761-2. Published online 13 August 2025; version-of-record issue June 2026.

Tong Chen, Zachary Quinn, Yinuo Zhang, and Pranam Chatterjee. moPPIt-v3: Motif-specific peptides generated via multi-objective-guided discrete flow matching. NeurIPS 2025 Workshop on

Structured Probabilistic Inference and Generative Modeling, 2025b. URL https://neurips.cc/virtual/2025/125414. Workshop poster; OpenReview: tHMtg5eDI6.

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.

Wenze Ding. BOND-PEP: topology-conditioned bipartite alignment for evidence-grounded peptide binder generation. *bioRxiv*, 2026. doi: 10.64898/2026.02.18.706554. URL https://doi.org/10.64898/2026.02.18.706554. Preprint posted 18 February 2026.

Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021. doi: 10.1101/2021.10.04.463034. URL https://doi.org/10.1101/2021.10.04.463034. Preprint posted 4 October 2021; not a peer-reviewed journal article.

Junxiong Huang, Weikang Li, Bin Xiao, Chunqing Zhao, Hancheng Zheng, Yingrui Li, and Jun Wang. PepCA: Unveiling protein-peptide interaction sites with a multi-input neural network model. *iScience*, 27(10):110850, 2024. doi: 10.1016/j.isci.2024.110850. URL https://doi.org/10.1016/j.isci.2024.110850.

Xiangzhe Kong, Yinjun Jia, Wenbing Huang, and Yang Liu. Full-atom peptide design with geometric latent diffusion. In *Advances in Neural Information Processing Systems*, volume 37, pp. 74808–74839, 2024. doi: 10.52202/079017-2379. URL https://doi.org/10.52202/079017-2379.

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 *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 27615–27640. PMLR, 2024. URL https://proceedings.mlr.press/v235/li24o.html.

Jiahan Li, Tong Chen, Shitong Luo, Chaoran Cheng, Jiaqi Guan, Ruihan Guo, Sheng Wang, Ge Liu, Jian Peng, and Jianzhu Ma. Hotspot-driven peptide design via multi-fragment autoregressive extension. In *International Conference on Learning Representations*, pp. 9282–9314, 2025a. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/1a0672689a693e0764f93f900488b3d9-Abstract-Conference.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, 2025b. doi: 10.1038/s42004-025-01601-3. URL https://doi.org/10.1038/s42004-025-01601-3.

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.

Pedro M. Martins, Lucianna H. Santos, Diego Mariano, Felippe C. Queiroz, Luana L. Bastos, Isabela de S. Gomes, Pedro H. C. Fischer, Rafael E. O. Rocha, Sabrina A. Silveira, Leonardo H. F. de Lima, Mariana T. Q. de Magalhães, Maria G. A. Oliveira, and Raquel C. de Melo-Minardi. Propedia: a database for protein–peptide identification based on a hybrid clustering algorithm. *BMC Bioinformatics*, 22(1):1, 2021. doi: 10.1186/s12859-020-03881-z. URL https://doi.org/10.1186/s12859-020-03881-z.

Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: making protein folding accessible to all. *Nature Methods*, 19(6): 679–682, 2022. doi: 10.1038/s41592-022-01488-1. URL https://doi.org/10.1038/s41592-022-01488-1.

Minjie Mou, Mingkun Lu, Zhimeng Zhou, Yanlin Ren, Xinyuan Yu, Ziqi Pan, Yuan Zhou, Hao Yang, Lingyan Zheng, Shukai Gu, Yang Zhang, Wei Hu, Fengcheng Li, Haibin Dai, and Feng Zhu. Accurate identification of protein binding sites for all drug modalities using ALLSites. *Advanced Science*, 13(10):e16530, 2026. doi: 10.1002/advs.202516530. URL https://doi.org/10.1002/advs.202516530.

Markus Muttenthaler, Glenn F. King, David J. Adams, and Paul F. Alewood. Trends in peptide drug discovery. *Nature Reviews Drug Discovery*, 20(4):309–325, 2021. doi: 10.1038/s41573-020-00135-8. URL https://doi.org/10.1038/s41573-020-00135-8.

Ngoc-Quang Nguyen, Jaeyoon Jung, Seijung Kim, Sunkyu Kim, and Jaewoo Kang. PepTri: Tri-guided all-atom diffusion for peptide design via physics, evolution, and mutual information. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=yQlTgHo1um.

Suzanne P. van Wier and Andrew M. Beekman. Peptide design to control protein–protein interactions. *Chemical Society Reviews*, 54(4):1684–1698, 2025. doi: 10.1039/D4CS00243A. URL https://doi.org/10.1039/D4CS00243A.

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. URL https://doi.org/10.1021/acs.jcim.4c00975.

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.

Jun Zhang, Yangyang Zhou, Tiantian Zhu, and Zexuan Zhu. PepCCD: A contrastive conditioned diffusion framework for target-specific peptide generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(33):28256–28264, 2026. doi: 10.1609/aaai.v40i33.40053. URL https://doi.org/10.1609/aaai.v40i33.40053.

A Implementation and Reproducibility

This appendix records the implementation settings and dataset statistics.

A.1 Two-stage training configuration

Both stages use the ESM-2 T33 (650M) backbone, AdamW with weight decay 0.01, and a maximum sequence length of 552. The stage-specific settings are summarized separately in Tables 6 and 7.

Table 6: Stage 1: site warm-up hyperparameters for SitePepGen.

Hyperparameter	**Value**
Objective	Site binary cross-entropy
Loss weights	$\lambda_{\mathrm{MLM}} = 0$, $\lambda_{\mathrm{site}} = 1$
Positive-class weight	$\alpha = 4$
Trainable components	Site head and ESM-2 layers 29–32
Training epochs	20
Batch size (train / evaluation)	8 / 16
Learning rate	Site head 3×10^{-4}; backbone 2×10^{-6}

Table 7: Stage 2: joint-training hyperparameters for SitePepGen.

Hyperparameter	**Value**
Objective	MLM + site binary cross-entropy
Loss weights	$\lambda_{\mathrm{MLM}} = \lambda_{\mathrm{site}} = 1$
Positive-class weight	$\alpha = 3$
Trainable components	All model parameters
Training epochs	5
Batch size (train / evaluation)	4 / 16
Peak learning rate	7.98×10^{-4}
Gradient accumulation steps	2 (effective batch size 8)
Learning-rate schedule	Cosine decay; 0.1 warm-up ratio

Stage 2 is initialized from the final Stage-1 checkpoint. The resulting checkpoint is frozen for candidate generation, site prediction, ranking, and X imputation; no additional selector adaptation is performed.

A.2 Candidate generation, selection, and imputation

The fixed inference settings are summarized below; the full procedure is given in Section 3.3 and Algorithm 1.

Table 8: Fixed inference settings for SitePepGen.

Parameter	**Setting**
Candidate pool	One initial candidate + 64 additional proposals
Top-k	3
Temperature	1.0
Site threshold	0.5 (after sigmoid)
Ranking score	$U(c) = Z_{\mathrm{compat}}(c) + Z_{\mathrm{LM}}(c)$
Compatibility weight γ	$1/4$
Selection safeguards	PPL ratio ≤ 1.05; composition $L_1 \leq 0.75$

For reproducibility, composition profiles are normalized over canonical residues, whereas peptide hydrophobic and charge fractions use the requested length L; an unresolved X therefore contributes

to the length but not to a property numerator. Hydrophobic residues are A, I, L, M, F, W, V, and Y; positive and negative residues are K/R and D/E, respectively.

A.3 Main inference algorithm

Algorithm 1 SitePepGen inference

Require: Receptor sequence r, peptide length L, frozen PG and SP with parameters θ
Ensure: Final peptide $c^\star \in \mathcal{A}_{20}^L$

$x \leftarrow (r, [\mathrm{MASK}]^L)$ {masked model input}
$c_0 \leftarrow \mathrm{SampleTop3}(P_\theta, x)$ {initial candidate}
$\mathcal{C} \leftarrow \mathrm{Unique}\big(c_0, \mathrm{SampleTop3}(P_\theta, x; 64)\big)$
$q_j \leftarrow \mathrm{SP}(x)_j$ for $j = 1, \ldots, M$ {receptor-site probabilities}
Derive site weights from the thresholded q_j.
Compute $S_{\mathrm{compat}}(c)$ and $S_{\mathrm{LM}}(c)$ for each $c \in \mathcal{C}$.
$U(c) \leftarrow Z_{\mathrm{compat}}(c) + Z_{\mathrm{LM}}(c)$ {standardize within $\mathcal{C}$}
$c^\star \leftarrow$ top-ranked candidate if safeguards relative to c_0 pass; otherwise c_0.
if $c^\star$ contains X **then**
  $c^\star \leftarrow \mathrm{ImputeX}(c^\star, P_\theta, x)$ {one-pass imputation}
end if
return $c^\star$

Here c_0 is the first generated peptide, retained as a fallback if the top-ranked candidate fails the selection safeguards; it is not a separate baseline model. The additional 64 proposals are sampled at temperature 1, and duplicate sequences are removed. Candidates are ranked by decreasing $U(c)$, then $S_{\mathrm{LM}}(c)$, with insertion order and sequence used to break remaining ties. $\mathrm{ImputeX}$ replaces only the original X positions: it tries a left-to-right canonical Top-1 pass, then the all-masked Top-1 completion and single-position Top-2/3 alternatives. The first imputation satisfying the fixed pseudo-perplexity and composition safeguards is returned; otherwise the all-masked completion is used.

A.4 Computational profile and exact structural software configuration

Table 9: Sequence-stage computational profile. GFLOPs are measured for one forward pass at the median length $T = 249$. Training time is measured per optimizer step, and test time is measured per query for sequence generation only; both exclude external ColabFold evaluation.

Method	Parameters	GFLOPs	Training time	Test time
PepMLM	652.36M	334.66	0.8449 s	0.0695 s
SitePepGen Phase 1	652.36M	334.66	0.3223 s	—
SitePepGen Phase 2	652.36M	334.66	0.8642 s	0.0960 s

Measurement details. Parameter totals count unique parameter objects, so tied embedding and decoder weights are counted once. The exact totals are 652,356,534 for the MLM-only model and 652,357,815 for the MLM+Site model; the site head adds 1,281 parameters. The GFLOPs value is measured on an NVIDIA RTX A6000 with PyTorch 2.0.1, CUDA 11.8, FP32, and batch size 1 using torch.profiler for a single backend forward. It excludes tokenization, sampling, ranking logic, infilling, and ColabFold. Training times are reported per optimizer step. Test times are reported per query for generation only; Phase 1 has no separate test-time result.

For post-generation structural evaluation, we used LocalColabFold 1.5.2. The MSA search used the MMseqs2 UniRef environmental mode (mmseqs2_uniref_env) together with the paired/unpaired mode (unpaired_paired). Templates, dropout, and relaxation were disabled; all other settings follow Section 4.3.

A.5 Dataset construction and statistics

We construct the dataset from protein–peptide complexes associated with PepNN, Propedia, and the Protein Data Bank (Abdin et al., 2022; Martins et al., 2021; Berman et al., 2000). Receptor

and peptide sequences are matched to their corresponding PDB chains before residue-level receptor binding-site labels are extracted. A receptor residue is positive when any of its heavy atoms lies within 5 Å of a peptide heavy atom.

The locked split contains 9,327 training pairs, 500 validation pairs, and 202 test queries. Table 10 summarizes the pair counts, unique receptor counts, and mean sequence and site statistics for each split. Unique receptors are counted by exact receptor-sequence identity, and the site fraction is first computed per receptor and then averaged within each split.

Table 10: Statistics of the locked receptor–peptide split. Lengths and site counts are means over rows.

Split	Pairs	Unique receptors	Receptor length	Peptide length	Site count	Site fraction
Train	9,327	6,796	223.07	14.89	20.41	10.92%
Validation	500	480	222.91	15.77	19.79	10.54%
Test	202	198	232.33	14.49	20.43	10.33%

A.6 Evaluation metric calculations

Receptor binding-site recovery. For each structurally evaluable query i, let R_i be the set of receptor binding-site residues in the reference complex and G_i the corresponding set in the predicted complex of the generated peptide. Both sets use the 5 Å heavy-atom criterion described in Section 4.3. We define

$$TP_i = |R_i \cap G_i|, \tag{7}$$
$$FP_i = |G_i \setminus R_i|, \tag{8}$$
$$FN_i = |R_i \setminus G_i|. \tag{9}$$

The query-level metrics are

$$J_i = \frac{TP_i}{TP_i + FP_i + FN_i}, \tag{10}$$
$$P_i = \frac{TP_i}{TP_i + FP_i}, \tag{11}$$
$$R_i^{\text{score}} = \frac{TP_i}{TP_i + FN_i}, \tag{12}$$
$$F1_i = \frac{2TP_i}{2TP_i + FP_i + FN_i}. \tag{13}$$

A metric is set to zero when its denominator is zero. For each method, the macro score is the arithmetic mean of its query-level scores over structurally evaluable queries.

Sequence diagnostics. For a generated peptide p_i of length L_i, pseudo-perplexity is computed by masking each peptide position in turn while keeping the receptor and the other peptide residues fixed. Let $\tilde{p}_i^{(\ell)}$ denote the peptide with position ℓ masked, and let P_{score} denote the same frozen scoring model used for all methods. Then

$$\text{PPL}(p_i \mid r_i) = \exp\left[-\frac{1}{L_i}\sum_{\ell=1}^{L_i} \log P_{\text{score}}\big(p_{i\ell} \mid r_i, \tilde{p}_i^{(\ell)}\big)\right]. \tag{14}$$

Amino-acid composition is calculated by pooling residues across the sequences in each group. For a canonical amino acid $a \in \mathcal{A}_{20}$, its percentage is

$$100\,\frac{\sum_i \sum_{\ell=1}^{L_i} \mathbf{1}[p_{i\ell} = a]}{\sum_i \sum_{\ell=1}^{L_i} \mathbf{1}[p_{i\ell} \in \mathcal{A}_{20}]}. \tag{15}$$

The same calculation is applied to the reference test peptides.

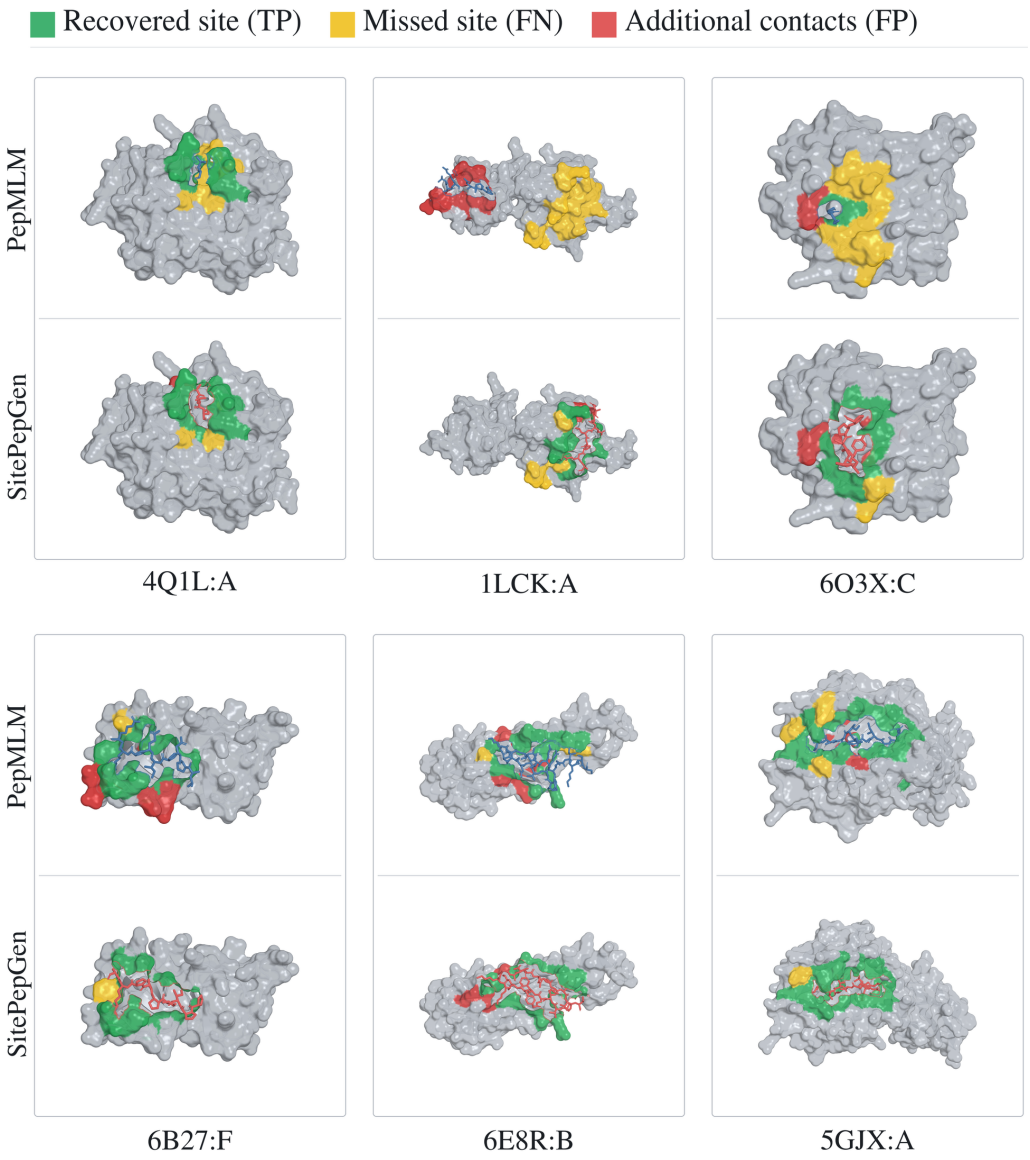
Predicted complex confidence. The ipTM value is taken directly from AlphaFold2-Multimer-v3. For the comparison with reference complexes, the reported fraction of generated complexes whose ipTM is at least as high as the corresponding reference ipTM is

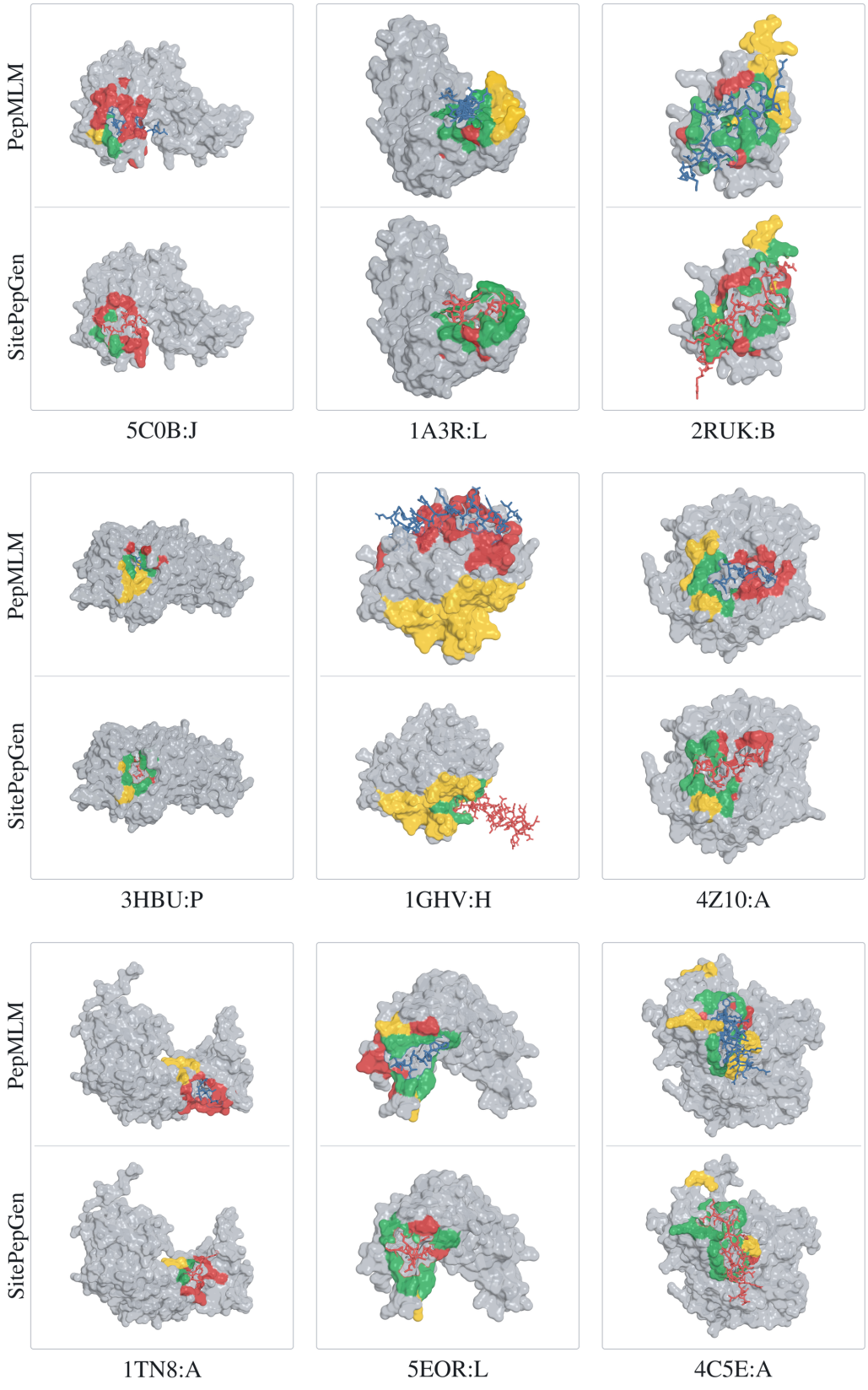
$$\frac{1}{|\mathcal{Q}_{\mathrm{ipTM}}|} \sum_{i \in \mathcal{Q}_{\mathrm{ipTM}}} \mathbf{1}\left[\mathrm{ipTM}_i^{\mathrm{gen}} \geq \mathrm{ipTM}_i^{\mathrm{ref}}\right], \tag{16}$$

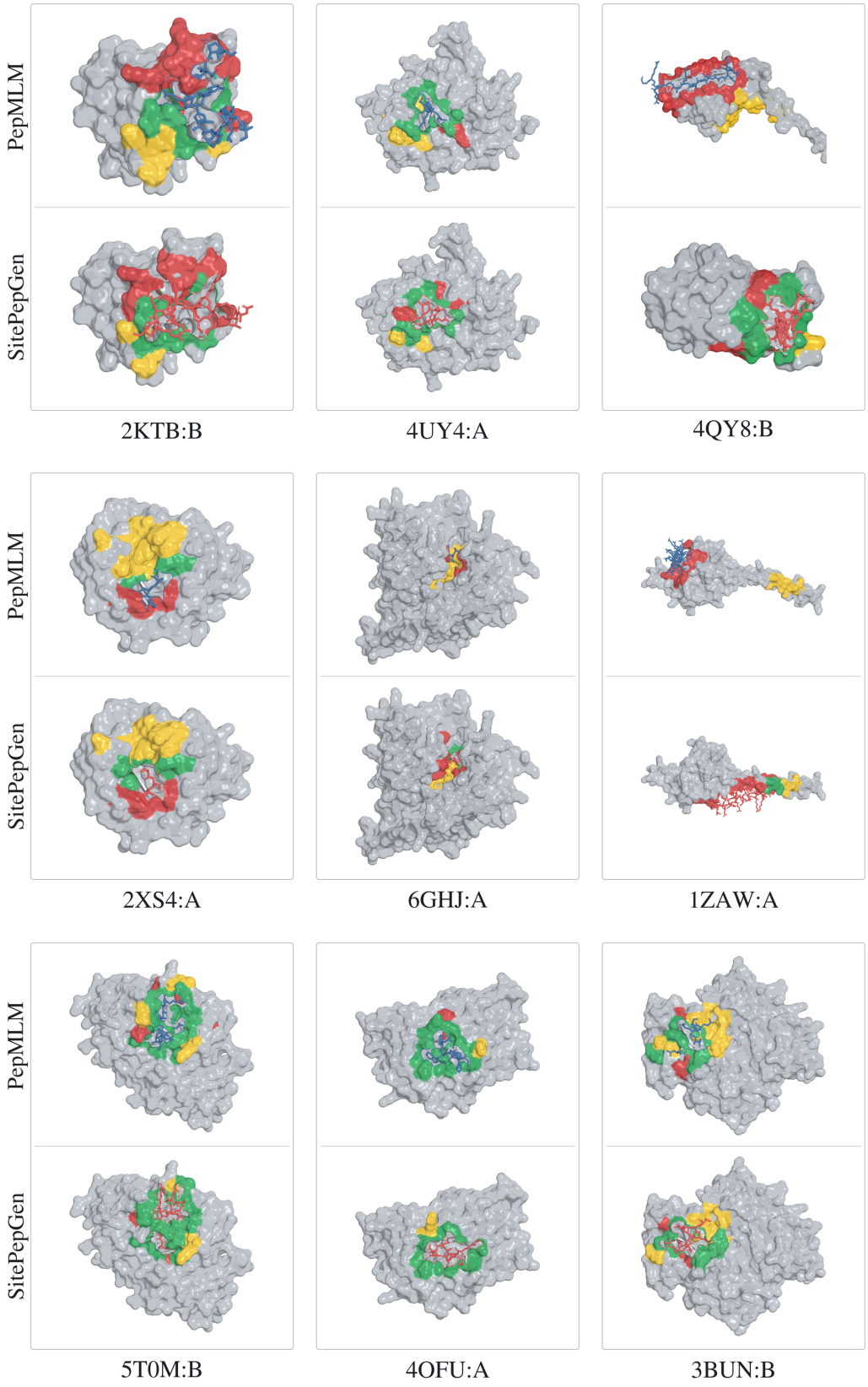
where $\mathcal{Q}_{\mathrm{ipTM}}$ contains queries with both values available.

A.7 Additional qualitative case studies

We show 24 selected paired cases in eight complete three-case rows. Within each case, PepMLM is shown in the upper square panel and SitePepGen in the lower square panel, with the receptor PDB-chain identifier below the pair. Green, yellow, and red indicate receptor residues assigned to the binding site in both the predicted and reference complexes (TP), only the reference complex (FN), and only the predicted complex (FP), respectively.




PepMLM
SitePepGen
5C0B:J
1A3R:L
2RUK:B
PepMLM
SitePepGen
3HBU:P
1GHV:H
4Z10:A
PepMLM
SitePepGen
1TN8:A
5EOR:L
4C5E:A


PepMLM
SitePepGen
2KTB:B
4UY4:A
4QY8:B
PepMLM
SitePepGen
2XS4:A
6GHJ:A
1ZAW:A
PepMLM
SitePepGen
5T0M:B
4OFU:A
3BUN:B