TRANSFERRING PEPTIDE–TARGET INTERACTION PATTERNS TO SMALL MOLECULES BY JOINT GUIDANCE FOR DIFFUSION AND FLOW MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Generating small molecules to target protein-protein interactions (PPIs) remains one of the most challenging yet high-reward frontiers in drug discovery. Unlike traditional small-molecule drugs that bind to deep, well-defined pockets, PPI interfaces often comprise broad, shallow binding surfaces. A natural strategy is to leverage known peptide–target interactions, as peptide ligands provide rich interface-level constraints. However, a major algorithmic challenge is that rigid atom-level conditioning on peptide structures can over-constrain the generative space and exclude valid small-molecule solutions. To address this challenge, we present PIFT (Peptide InterFace Transfer), a novel framework that transfers peptide–target interaction patterns to small molecules through joint sampling-time guidance of masked discrete diffusion and continuous flow matching. PIFT first maps peptide–target and small-molecule–target interfaces into a shared, target-conditioned interaction space, capturing transferable interface information across molecular modalities. The resulting peptide-derived reference then jointly steers building-block logits and 3D coordinates toward peptide-compatible molecules while preserving a strong adapted molecular prior. Extensive experiments across diverse PPI targets show that the learned representation captures cross-modal interaction compatibility, while peptide guidance improves interface recovery and achieves strong docking performance and favorable molecular properties.

1 INTRODUCTION

Protein–protein interactions (PPIs) regulate diverse cellular processes and are broadly implicated in human disease, making them an important class of therapeutic targets (Greenblatt et al., 2024; Nada et al., 2024). Their interfaces, however, are typically broad, shallow, and weakly pocketed, with extended and heterogeneous surface geometries that make effective small-molecule engagement particularly challenging (Arkin & Wells, 2004; Jin et al., 2014). Recent structure-based generative models have substantially expanded the design space by conditioning molecular identity and three-dimensional geometry on protein structural context (Peng et al., 2022; Guan et al., 2023), while newer synthesis-aware formulations additionally couple molecular generation to chemically feasible construction pathways (Cretu et al., 2025; Seo et al., 2025; Shen et al., 2025). These models can further incorporate reference-ligand information (Huang et al., 2024) or explicit protein–ligand interaction constraints (Zhung et al., 2024), providing increasingly fine-grained control over generation. For PPIs, however, specifying an appropriate binding condition remains difficult: the relevant interface may extend beyond a compact ligand-binding site, leaving the desired interface constraint under-specified. This motivates PPI-specific methods that derive design cues directly from the target interaction interface.

A prominent class of PPI-specific methods derives molecular design constraints from interface hotspots. In particular, hotspot-derived pharmacophore conditioning represents a broad PPI interface through a compact spatial arrangement of selected energetically or functionally important features (Wells & McClendon, 2007; Sun et al., 2026). The effectiveness of such methods establishes the utility of interface-derived information for small-molecule design. Nevertheless, hotspot conditioning intentionally compresses the interface into a selected subset of spatial and chemical cues, limiting the extent to which broader multi-contact organization is retained.

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.

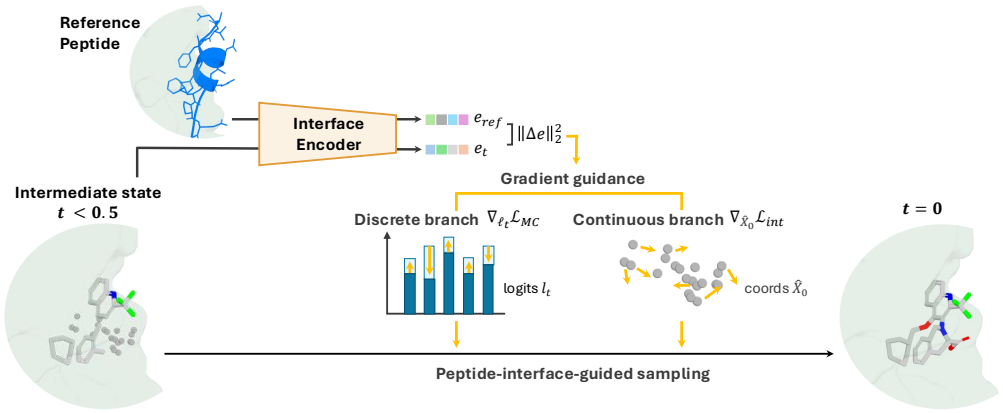


Figure 1: **Overview of PIFT molecular generation framework.** The interface encoder extracts representations from a reference peptide–target interface and an intermediate molecule–target state. The guidance objective derived from their embeddings jointly steers discrete building-block logits and continuous 3D coordinates during sampling.

An alternative is to draw design information from the interaction partner itself. Many PPIs are mediated by short peptide motifs, and bound peptides therefore provide structurally realized examples of how multiple contacts are organized across an extended target surface (Corbi-Verge et al., 2017; Pelay-Gimeno et al., 2015; van Wier & Beekman, 2025). Recent peptide-conditioned generative approaches encode peptide and protein structures at the atomic level, often augmented with residue-level context, and inject these representations into molecular denoisers through equivariant or attention-based conditioning mechanisms (He et al., 2025; Yue et al., 2026).

However, transferring peptide-defined interfaces into small-molecule generation still poses two challenges.

- **Structural mismatch across molecular modalities.** Peptides and small molecules realize binding through fundamentally different molecular organizations: peptides use residue-indexed backbone and side-chain architectures, whereas small molecules arrange heterogeneous functional groups on compact scaffolds. Relying on fine-grained atomic representations of peptide structure can over-constrain the space of valid small-molecule realizations. This structural mismatch poses a fundamental challenge for peptide-to-small-molecule translation.
- **Guidance across discrete chemistry and 3D geometry.** Once the peptide interface is transferred into a shared representation, using it to control generation remains nontrivial. Synthesis-aware molecular generation involves both discrete building-block assembly and continuous three-dimensional geometry. Because these two components respond to conditioning in fundamentally different ways, incorporating peptide-derived constraints without distorting the pretrained molecular distribution remains challenging.

To address these challenges, we introduce PIFT, formulating peptide-to-small-molecule generation as transferring peptide–target interaction patterns through joint control of discrete molecular identity and continuous 3D geometry. PIFT learns a shared target-conditioned interaction space, adapts a pretrained synthesis-aware generator toward a target-specific PPI-oriented prior, and uses the peptide interface as a sampling reference (Figure 1).

Our contributions are:

- We introduce a novel target-conditioned cross-modal representation that transfers peptide–target interaction patterns to small molecules through shared interaction types and three-dimensional geometry.
- We develop a joint peptide-reference guidance mechanism that applies a single interface objective to both masked discrete diffusion over building blocks and continuous flow matching over 3D geometry, enabling coordinated control of chemistry and pose.

- Across diverse PPI targets, we show that the learned representation captures cross-modal interaction compatibility and that PIFT effectively recovers the peptide interface while achieving strong docking performance and molecular properties.

2 Background and Related Work

2.1 Pharmacophore and Peptidomimetic Design for PPI Inhibitors

Peptidomimetic design replaces a peptide or protein interaction motif with a chemically distinct scaffold while preserving the key determinants required for binding. Pharmacophore-based approaches provide a natural abstraction for this transfer by representing interaction-relevant functional features and their spatial organization independently of the underlying molecular topology (Wells & McClendon, 2007; Voet & Zhang, 2012; Pelay-Gimeno et al., 2015; Qvit et al., 2017). This abstraction motivates our use of a shared pharmacophore representation to compare peptide and small-molecule interfaces under the same target context.

2.2 Discrete Synthesis-Aware Molecular Generation

Masked discrete diffusion. Let x_0 denote a categorical object represented over a vocabulary that includes an absorbing mask token. Discrete denoising diffusion progressively corrupts x_0 according to a cumulative transition matrix $\bar{Q}_t$ and trains a reverse model to predict the clean categorical state (Austin et al., 2021):

$$q(x_t \mid x_0) = \mathrm{Cat}(x_t; x_0\bar{Q}_t), \qquad p_\theta(x_0 \mid x_t, t). \tag{1}$$

With absorbing-mask corruption, generation proceeds by iteratively resolving masked variables. The reverse transition can be written in terms of the predicted clean-state posterior as

$$p_\theta(x_s \mid x_t) = \sum_{x_0} q(x_s \mid x_t, x_0) p_\theta(x_0 \mid x_t), \qquad s < t, \tag{2}$$

which exposes the clean-state prediction as a natural point for sampling-time categorical guidance. In synthesis-aware molecular generators, these categorical variables can represent building blocks and reaction choices, grounding generation in synthesizable chemical space.

2.3 Sampling-Time Guidance in Continuous and Discrete Spaces

Continuous gradient guidance. Sampling-time guidance controls a pretrained generative trajectory through an external objective without retraining the full model (Dhariwal & Nichol, 2021; Bansal et al., 2024). If a differentiable loss $\mathcal{L}(\hat{x}_0, c)$ is evaluated on the predicted clean state $\hat{x}_0 = \hat{x}_{0,\theta}(x_t)$, the reverse score can be modified as

$$\widetilde{s}_\theta(x_t, t; c) = s_\theta(x_t, t) - \lambda_t \nabla_{x_t} \mathcal{L}(\hat{x}_{0,\theta}(x_t), c). \tag{3}$$

This illustrates the general principle of differentiating a downstream objective through a denoised prediction; our geometric guidance applies the resulting correction directly to the predicted clean coordinates.

Guidance for categorical states. Discrete diffusion poses a distinct challenge for sampling-time guidance because its categorical states are not directly differentiable. GILC addresses this difficulty by building on straight-through Gumbel–Softmax (Jang et al., 2017) to obtain gradients through discrete samples and applying them as Jacobian-free corrections to clean-prediction logits (Dou et al., 2026). This provides the basis for our discrete guidance branch, which applies peptide-reference gradients to building-block logits during sampling.

2.4 Reward-Based Adaptation of Generative Priors

Reward tilting adapts a pretrained generative distribution toward high-reward regions while retaining support from the original prior. For masked discrete diffusion, Discrete Tilt Matching (DTM) realizes this principle by matching reward-tilted local unmasking posteriors at intermediate states, avoiding direct sequence-level likelihood optimization (Chen et al., 2026; Wang et al., 2025). We use DTM to construct the PPI-oriented molecular prior described in Section 3.3.1.

3 METHOD

3.1 PROBLEM FORMULATION

For a target protein with a bound reference peptide, let C_{pep} and T the denote the target-facing peptide and receptor interface representations, respectively. We generate a synthesizable molecule and pose $x = (G, \mathbf{X})$, where G denotes the reaction graph, including building-block identities and coupling reaction (and centers), and and $\mathbf{X} \in \mathbb{R}^{N\times 3}$ contains the corresponding 3D coordinates. The desired output should recover a target-facing interaction pattern compatible with that of the reference peptide.

3.2 TRANSFERRING PEPTIDE–TARGET INTERACTION PATTERNS TO SMALL MOLECULES

3.2.1 SHARED INTERFACE REPRESENTATION

Dataset construction. We construct the cross-modal interface corpus from peptide–protein and small-molecule–protein complexes in **PepBDB**, **CrossDocked2020**, **2P2Idb v2**, and **DLiP-PPI** (Wen et al., 2019; Francoeur et al., 2020; Basse et al., 2016; Ikeda et al., 2023). Complexes are aligned within a common target coordinate frame, and splits are defined by UniProt accession to prevent target leakage. Detailed processing and corpus statistics are provided in Appendix A.1.1.

Shared Representation. We represent peptide and small-molecule interfaces through target-facing interaction features extracted with ProLIF (Bouysset & Fiorucci, 2021). Peptide and small-molecule interfaces share the same eight-type vocabulary: hydrogen-bond donor and acceptor, aromatic, hydrophobic, cationic, anionic, halogen-bond, and metal-binding features. Each binder-side interface is represented as a typed three-dimensional point cloud $C = \{(\mathbf{x}_i, c_i, \mathbf{d}_i, v_i)\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbb{R}^3$ denotes the feature coordinate, c_i the interaction type, $\mathbf{d}_i$ an optional interaction direction, and $v_i \in \{0, 1\}$ a validity mask. The corresponding target cloud T uses complementary receptor-side interaction types associated with observed contacts, with each point placed at the closest participating target atom for its residue-pair/type combination.

3.2.2 TARGET-CONDITIONED CROSS-MODAL INTERFACE ENCODER.

Encoder architecture. Peptide and small-molecule interfaces are processed by a shared encoder. Each interface point is initialized from its interaction type and SE(3)-invariant geometric features, followed by pairwise message passing and masked pooling:

$$\mathbf{h}_i^{(0)} = f_{\text{point}}([\text{Emb}(c_i), \boldsymbol{\rho}_i])\, v_i, \tag{4}$$

$$\mathbf{h}_i^{(1)} = \text{LN}\left(\mathbf{h}_i^{(0)} + \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} f_{\text{msg}}\left(\mathbf{h}_i^{(0)}, \mathbf{h}_j^{(0)}, \boldsymbol{\rho}_{ij}\right)\right) v_i, \tag{5}$$

$$\mathbf{p} = \frac{\sum_i v_i \mathbf{h}_i^{(1)}}{\sum_i v_i}. \tag{6}$$

where $\mathcal{N}(i) = \{j : v_j = 1,\ \|\mathbf{x}_i - \mathbf{x}_j\|_2 < r_c\}$, and $\boldsymbol{\rho}_i$ and $\boldsymbol{\rho}_{ij}$ denote translation- and rotation-invariant geometric scalars. A separate target encoder produces $\mathbf{t} = E_{\text{target}}(T)$, which is mapped to FiLM (Perez et al., 2018) parameters and applied to the pooled representation before final projection:

$$\begin{aligned}(\boldsymbol{\gamma}, \boldsymbol{\beta}) &= f_{\text{FiLM}}(\mathbf{t}),\\ E_\phi(C, T) &= f_{\text{out}}\left(\boldsymbol{\gamma} \odot \mathbf{p} + \boldsymbol{\beta}\right) \in \mathbb{R}^{64}.\end{aligned} \tag{7}$$

The FiLM modulation is initialized at $\boldsymbol{\gamma} = \mathbf{1}$ and $\boldsymbol{\beta} = \mathbf{0}$, so that target conditioning initially acts as an identity transformation.

Phase I: modality-shared geometric pretraining. We first pretrain target-agnostic $E_\phi(C)$ on balanced peptide and small-molecule batches. Following the training scheme of PepMimic (Kong et al., 2026), each cloud yields two overlapping views C_a, C_b and a spatially disjoint view C_n:

$$\mathcal{L}_{\text{partial}} = \|E_\phi(C_a) - E_\phi(C_b)\|_2 + \text{ReLU}(m_{\text{p}} - \|E_\phi(C_a) - E_\phi(C_n)\|_2)\,. \tag{8}$$

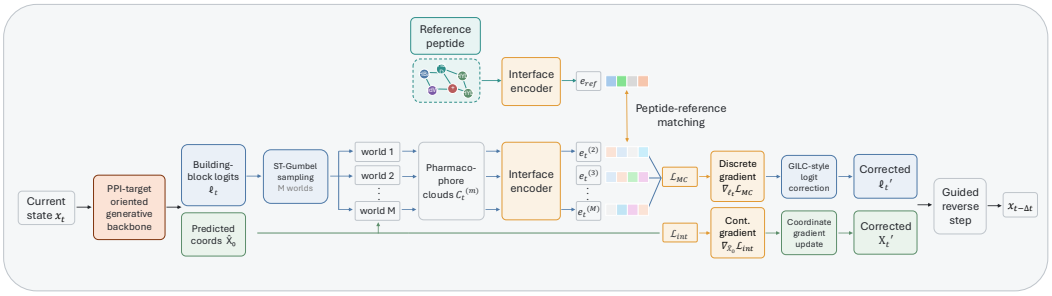


Figure 2: **Branch-specific mechanism of peptide-reference guidance.**

The objective brings overlapping views together while separating spatially disjoint views, providing a target-agnostic initialization for Phase II.

Phase II: target-conditioned cross-modal alignment. Starting from the Phase-I encoder, we introduce the target encoder and FiLM conditioning and train peptide queries with same-target molecular examples. All examples in a query group share target representation $\mathbf{t}_q$. For query q, positive p, and negative n,

$$\mathcal{L}_{\text{rank}} = \mathbb{E}_{q,p,n}\left[\text{softplus}\left(\frac{\|\mathbf{e}_q - \mathbf{e}_p\|_2 - \|\mathbf{e}_q - \mathbf{e}_n\|_2 + m_{\text{r}}}{\tau_{\text{r}}}\right)\right], \tag{9}$$

where $\mathbf{e}_u = E_\phi(C_u, T_q)$. We regularize the FiLM scale and discourage collapse of target representations:

$$\begin{gathered}\mathcal{L}_\gamma = \mathbb{E}\|\boldsymbol{\gamma} - \mathbf{1}\|_2^2, \qquad \mathcal{L}_{\text{div}} = \mathbb{E}_{k\neq l}\left[\text{ReLU}(\cos(\mathbf{t}_k, \mathbf{t}_l) - \tau_{\text{div}})^2\right], \\ \mathcal{L}_{\text{enc}} = \mathcal{L}_{\text{rank}} + \lambda_\gamma \mathcal{L}_\gamma + \lambda_{\text{div}} \mathcal{L}_{\text{div}}.\end{gathered} \tag{10}$$

The resulting space places peptide and small-molecule interfaces with compatible target-facing interactions closer together under the same target context.

3.3 Guiding Molecular Generation toward Peptide–Target Interfaces

3.3.1 PPI-Oriented Molecular Prior Adaptation

We use SynCoGen (Rekesh et al., 2026) as the synthesis-aware generative backbone and adapt its pretrained molecular prior toward a target-specific PPI-oriented distribution using Discrete Tilt Matching (DTM). Given the pretrained prior $p_0(x)$, DTM defines the reward-tilted distribution

$$p_R(x \mid T) = \frac{1}{Z_R(T)} p_0(x) \exp(\beta R_{\text{PPI}}(x, T)), \qquad Z_R(T) = \mathbb{E}_{p_0}\left[\exp(\beta R_{\text{PPI}}(x, T))\right]. \tag{11}$$

DTM then adapts the discrete denoiser toward this reward-tilted distribution through local unmasking-posterior matching. Details of the reward and optimization are provided in Appendix A.2.

3.3.2 Joint Guidance of Discrete Diffusion and Continuous Flow Matching

Shared interface objective. After target adaptation, the generator and interface encoder are frozen, and reference embedding $\mathbf{e}_{\text{ref}} = E_\phi(C_{\text{pep}}, T)$ is computed once. At reverse step t, the backbone predicts clean-state building-block logits $\boldsymbol{\ell}_t$ and a denoised coordinate estimate $\hat{\mathbf{X}}_0$. Let $\bar{E}_t$ denote the committed graph connectivity, which is held fixed during guidance. For an interface realization C, we define the shared peptide-reference objective $\mathcal{L}_{\text{int}}(C) = \|E_\phi(C, T) - \mathbf{e}_{\text{ref}}\|_2^2$. The discrete and continuous branches optimize this same objective through different differentiable realizations of molecular identity and geometry (Figure 2).

Discrete building-block guidance. For each unresolved building-block node, we construct M differentiable categorical realizations from the clean-state logits ℓ_t using straight-through Gumbel–Softmax (Jang et al., 2017):

$$\mathbf{y}_n^{(m),\mathrm{soft}} = \mathrm{softmax}\left(\frac{\ell_{t,n} + \boldsymbol{\eta}_n^{(m)}}{\tau_G}\right), \qquad \mathbf{y}_n^{(m),\mathrm{ST}} = \mathrm{sg}\left(\mathbf{y}_n^{(m),\mathrm{hard}} - \mathbf{y}_n^{(m),\mathrm{soft}}\right) + \mathbf{y}_n^{(m),\mathrm{soft}}, \tag{12}$$

where $\boldsymbol{\eta}_n^{(m)}$ is i.i.d. Gumbel noise and $\mathrm{sg}(\cdot)$ denotes stop-gradient. Committed building blocks are shared across worlds, while unresolved nodes may realize different candidate identities.

Each world induces the pharmacophore realization $C_t^{(m)} = \Phi\left(\mathbf{y}^{(m),\mathrm{ST}}, \bar{E}_t, \hat{\mathbf{X}}_0\right)$, yielding the Monte Carlo interface objective

$$\mathcal{L}_{\mathrm{MC}} = \frac{1}{M}\sum_{m=1}^{M}\left\|E_\phi(C_t^{(m)}, T) - \mathbf{e}_{\mathrm{ref}}\right\|_2^2. \tag{13}$$

Its gradient $\mathbf{g}_\ell = \nabla_{\ell_t}\mathcal{L}_{\mathrm{MC}}$ is converted into a GILC-style clean-logit correction (Dou et al., 2026). The gradient is restricted to unresolved nodes with valid reaction-supported realizations, then normalized, clipped, and smoothed.

We apply the guidance only during the later reverse steps using the schedule

$$r_{\mathrm{disc}}(t) = \left[\max\left(0, \frac{t_{\mathrm{start}} - t}{t_{\mathrm{start}}}\right)\right]^{p_{\mathrm{disc}}}. \tag{14}$$

With a fixed maximum discrete guidance scale $\gamma_{\max}$, the corrected clean logits are

$$\boldsymbol{\ell}_{t,n} \leftarrow \boldsymbol{\ell}_{t,n} - \gamma_{\max} r_{\mathrm{disc}}(t)\hat{\mathbf{g}}_{\ell,n}. \tag{15}$$

Continuous geometric guidance. Because the interface cloud is differentiable with respect to $\hat{\mathbf{X}}_0$, we directly compute $\mathbf{g}_X = \nabla_{\hat{\mathbf{X}}_0}\mathcal{L}_{\mathrm{int}}(C_t)$. After removing the global translation component and applying normalization and clipping, we update

$$\hat{\mathbf{X}}_0 \leftarrow \hat{\mathbf{X}}_0 - t\,\gamma_C(t)\hat{\mathbf{g}}_X, \qquad \gamma_C(t) = \gamma_{\max}\left[\max\left(0, \frac{t_{\mathrm{start}} - t}{t_{\mathrm{start}}}\right)\right]^{p_C}. \tag{16}$$

This applies the same peptide-reference objective directly to 3D geometry.

We apply RMS normalization, clipping, and adaptive moment smoothing to the guidance gradients; implementation details are provided in Appendix A.3.2.

4 Experiments

4.1 Experimental Setup

Targets and sampling protocol. We evaluate our method and external baselines on 10 predefined PPI targets spanning distinct interface geometries and target families. For the guidance ablation, we compare PIFT with the same target-adapted molecular prior without peptide-reference guidance (Off), holding target, prior checkpoint, sampling budget, and random seeds fixed whenever supported.

Baselines. We compare against representative alternatives covering structure-based 3D generation, synthesis-aware molecular generation, and peptide/PPI-guided design. **TargetDiff**, **MolCRAFT**, and **FLOWR** provide strong structure-conditioned diffusion or flow-based baselines (Guan et al., 2023; Qu et al., 2024; Cremer et al., 2026); **CGFlow**, **SynCoGen**, and **SynGA** represent synthesis-aware or search-based generation (Shen et al., 2025; Rekesh et al., 2026; Lo et al., 2026). CGFlow is evaluated in a target-specific setting, with a separate model trained for each target. We further include **Hot2Mol** as a hotspot-guided PPI inhibitor generator and **Peptide2Mol** as a peptide-conditioned peptidomimetic diffusion model (Sun et al., 2026; He et al., 2025). Architectural details are provided in Appendix B.1.

Metrics. While peptidomimetic design emphasizes preservation of key binding determinants across distinct scaffolds (Pelay-Gimeno et al., 2015), standard docking and molecular-quality metrics do not assess recovery of the reference peptide's target-facing interaction pattern. We therefore introduce **Interface Agreement (IFA)**, which compares a reference peptide P and molecule L through residue-wise target interaction profiles. For binder $B \in \{P, L\}$, target residue r, and ProLIF interaction type t, we define

$$\mathbf{u}_B(r,t) = \begin{cases} \exp\left[-d^B_{\min}(r,t)/(3.0\,\text{Å})\right], & \text{if an interaction of type } t \text{ is present,} \\ 0, & \text{otherwise,} \end{cases}$$

and compute

$$\text{IFA}(P,L) = \sum_{r:\, \|\mathbf{u}_P(r)\|+\|\mathbf{u}_L(r)\|>0} \cos(\mathbf{u}_P(r), \mathbf{u}_L(r))\,.$$

Here $d^B_{\min}(r,t)$ denotes the closest interaction of type t between binder B and residue r; repeated interactions retain the minimum distance. IFA is validated against target-residue overlap on held-out pairs, showing strong agreement and supporting its reliability as an interface-similarity metric (Appendix C.1).

We report the **Vina docking score (kcal/mol)** for docking quality (Eberhardt et al., 2021), **Validity (%)** as the proportion of generated molecules identified as chemically valid by RDKit, **Uniqueness (%)** as the proportion of unique generated molecules, and **Diversity** as the mean pairwise Tanimoto distance between ECFP4 fingerprints (Rogers & Hahn, 2010). We additionally report **AiZynth.** as the AiZynthFinder success rate, and **PoseCheck** as the summed protein–ligand interaction count in the docked pose. To ensure a fair aggregate comparison, PoseCheck results are reported only for models with successful evaluation on all 10 targets.

4.2 Transferring Peptide-Target Interaction Patterns to Small Molecules

We evaluate whether the learned cross-modal interface encoder captures interaction-pattern similarity by relating peptide–molecule embedding distance to target-residue overlap and IFA. Across held-out pairs, stronger interface agreement is associated with smaller embedding distances (Figure 3).

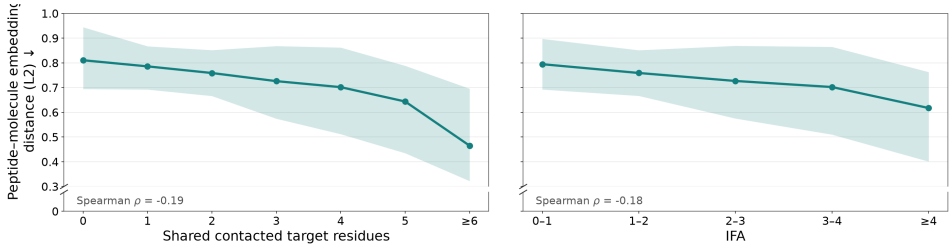


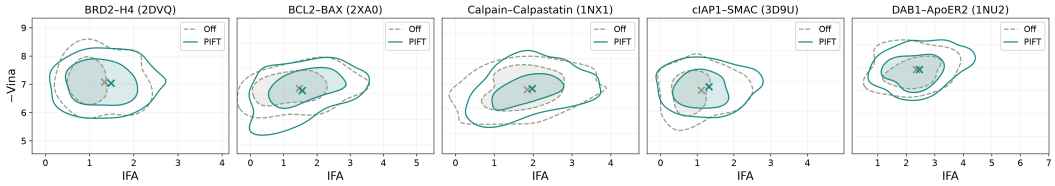
Figure 3: **Peptide–molecule embedding distance across target-residue overlap and IFA.** *Left:* Embedding distance decreases as shared contacted target residues increase, clearly distinguishing overlap levels. *Right:* Higher IFA shows the same trend. Lines indicate the median and shaded regions the interquartile range (Q1–Q3).

4.3 Generating PPI Inhibitor with Peptide-Target Interface Guidance

Overall population shift. We first assess whether peptide-reference guidance shifts generation toward molecules that better recover the desired interface while preserving the binding quality of the adapted prior. Relative to Off, guidance largely preserves the Vina distribution while extending sampling into higher-IFA regions, enabling discovery of molecules with stronger peptide-interface agreement (Figure 4a).

Joint guidance ablation. To evaluate the effect of joint guidance, we ablate the discrete and continuous geometric branches (Figure 4b). Coordinate guidance achieves the strongest single-branch docking score and the highest target-residue coverage, whereas discrete guidance yields higher interface precision despite contacting fewer residues. These differences suggest complementary roles:

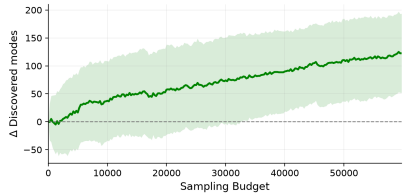
(a)



(b)

Guidance	IFA ↑	Precision ↑	Res. count	Vina ↓	Success ↑
Off	1.975	0.448	3.360	-7.662	25.3%
Coordinate	2.016	0.444	3.439	**-7.708**	26.5%
Discrete	2.016	0.457	3.425	-7.691	24.7%
Joint	**2.049**	**0.460**	3.385	-7.703	**29.3%**

(c)



(d)

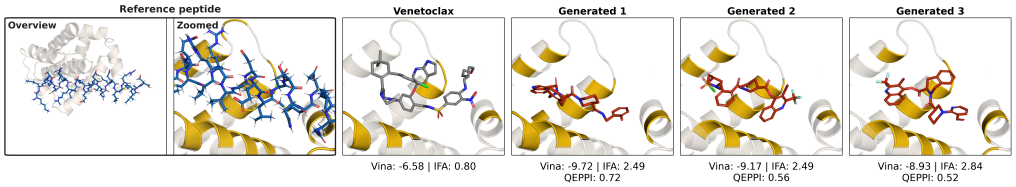


Figure 4: **Peptide-reference joint guidance. (a) PIFT and Off populations in IFA–−Vina space.** Shared density regions show that guidance largely preserves the distribution supported by the DTM-adapted prior, while guided contours extend toward higher interface agreement. Full results for the 10 PPI targets are provided in Appendix C.4.**(b) Joint guidance ablation.** Precision measures agreement of generated and reference (residue, interaction type) contacts; success requires both IFA above and Vina below the target-specific Off median. **(c) Fixed-budget discovery.** The curve reports the cumulative number of additional qualified modes discovered by PIFT relative to Off; the line is the median and shading the approximate 95% interval. **(d) 3D illustration of BCL2–BAX reference peptide, Venetoclax, and guided molecules** Generated molecules demonstrate chemically distinct realizations of the reference interface.

coordinate updates improve the spatial realization of the molecule at the interface, while discrete updates favor building blocks with more reference-compatible interaction functionality. Joint guidance combines both effects, achieving the highest IFA, precision, and success rate, while retaining near-best docking quality without unnecessarily expanding residue coverage.

Fixed-budget discovery. We evaluate discovery under matched sampling budgets for each target, defining a qualified mode as IFA > 1.5, QEPPI > 0.5, Vina < -6 kcal/mol, and Tanimoto similarity < 0.5 to any previously discovered mode.[1]. Figure 4c reports the number of qualified modes across targets relative to Off. The increase shows that peptide-reference guidance more effectively discovers molecules that jointly satisfy interface compatibility, docking quality, and diversity criteria.

Reference specificity. We evaluate reference specificity on MDM2 using two experimentally resolved peptide binders, 1YCR and 3LNJ, whose target-contact patterns partially overlap but remain distinguishable. The results show clear reference-dependent guidance, indicating that the model responds to the selected peptide interaction pattern rather than simply favoring generic MDM2-binding features. Detailed results are provided in Appendix C.3.

4.4 Molecular Quality and Binding Performance

External baseline comparison. Across individual PPI targets, PIFT consistently outperforms or remains competitive with diverse generative baselines, achieving the best mean and median Vina scores (-7.68/-7.83, Table 2) and the highest aggregate IFA (1.993, Table 3). Notably, PIFT shows

[1]For IL2, we use Vina < -4 kcal/mol to identify a comparable number of discovered modes relative to the other targets.

a clear advantage over diffusion- and flow-based baselines that degrade sharply on difficult PPI targets. Both diffusion's stochastic denoising and flow matching's deterministic transport rely on a localized ligand-sized region defined by receptor geometry, making their conditioning less reliable when no such cavity is well defined. This vulnerability is apparent on BCL2–BAX, where TargetDiff and Peptide2Mol fail to generate usable molecules and FLOWR becomes structurally degenerate, while PIFT reaches the best Vina score (-8.22) among all baselines. We attribute this robustness to conditioning on the peptide-defined cross-modal interaction pattern rather than relying solely on receptor geometry. Beyond docking, our model achieves the highest aggregate IFA and PoseCheck interaction count, while retaining perfect uniqueness and a high fraction of chemically valid molecules.

Table 2: **Vina docking performance.** Average Vina docking score for each targets, and the average and median values across 10 PPI targets. Results for remaining 5 targets are reported in Appendix C.4. Best results are in bold.

Category	Model	Average Vina Docking Score (kcal/mol, ↓)						
		BRD2–H4	BCL2–BAX	Calpain–Calpastatin	cIAP1–SMAC	DAB1–ApoER2	Avg.	Med.
Autoregressive	Hot2Mol	-7.00 (± 0.04)	-7.61 (± 0.05)	-6.56 (± 0.03)	-6.59 (± 0.05)	-7.35 (± 0.02)	-7.01	-7.06
Search-based	SynGA	-6.47 (± 0.09)	-7.33 (± 0.05)	-6.50 (± 0.03)	-6.04 (± 0.09)	-6.78 (± 0.03)	-6.62	-6.64
GFlowNet-based	CGFlow	**-7.20** (± 0.05)	-7.93 (± 0.03)	-6.91 (± 0.03)	-6.38 (± 0.03)	-6.99 (± 0.05)	-7.03	-7.10
Diffusion / Flow-based	TargetDiff	-5.30 (± 0.04)	–	-6.74 (± 0.03)	-3.50 (± 0.03)	-6.02 (± 0.08)	-6.27	-6.82
	MolCRAFT	-5.02 (± 0.02)	-6.84 (± 0.03)	-7.50 (± 0.04)	-3.90 (± 0.04)	-5.32 (± 0.01)	-5.81	-6.25
	Flowr	-4.00 (± 0.19)	-1.25 (± 0.00)[†]	-5.00 (± 0.08)	-6.39 (± 0.05)	-6.04 (± 0.21)	-5.74	-5.91
	SynCoGen	-6.68 (± 0.08)	-7.70 (± 0.04)	-7.00 (± 0.05)	-6.33 (± 0.02)	-7.05 (± 0.08)	-6.92	-7.00
	Peptide2Mol	-5.68 (± 0.54)	–	-6.18 (± 0.33)	-5.12 (± 0.23)	-4.90 (± 0.07)	-5.51	-5.59
	PIFT	-6.99 (± 0.04)	**-8.22** (± 0.06)	**-7.75** (± 0.06)	**-6.78** (± 0.05)	**-7.92** (± 0.04)	**-7.68**	**-7.83**

Table 3: **Molecular-generation quality.** Best results are in bold.

Category	Model	IFA ↑	Validity ↑	AiZynth. ↑	Uniqueness ↑	Diversity ↑	PoseCheck ↑
Autoregressive	Hot2Mol	1.347	0.975	0.529	**1.000**	0.819	10.485
Search-based	SynGA	1.329	**1.000**	0.258	**1.000**	0.878	11.071
GFlowNet-based	CGFlow	1.182	0.978	0.575	0.972	0.828	9.193
Diffusion / Flow-based	TargetDiff	1.227	0.571	0.106	**1.000**	0.891	—
	MolCRAFT	1.420	0.864	0.226	0.976	0.854	11.340
	Flowr	1.290	0.909	0.023	**1.000**	0.842	—
	SynCoGen	1.384	0.957	**0.643**	**1.000**	0.804	10.648
	Peptide2Mol	1.068	0.088	0.011	**1.000**	**0.893**	—
	PIFT	**1.993**	0.947	<u>0.612</u>	**1.000**	0.709	**12.351**

5 Conclusion

In this work, we introduced PIFT, a novel peptide-guided framework for small-molecule PPI inhibitor generation. PIFT effectively represents peptide-target and small-molecule-target interaction patterns in a shared interface space, and steers the molecular prior toward peptide-compatible regions.

Across diverse and challenging PPI targets, PIFT identifies peptide-compatible molecules with strong docking performance while retaining predicted synthesizability and favorable molecular properties. These results demonstrate PIFT's ability to effectively discover diverse successful molecular candidates for PPI interfaces where conventional 3d molecular generation models struggle.

PIFT thus offers a new route toward the long-standing challenge of small-molecule PPI inhibitor design. Beyond peptide-to-small-molecule transfer, future work should extend this principle to interaction-guided generation across diverse biomolecular modalities.

AI USE STATEMENT

In this work, we used generative AI tools to polish the writing by refining grammar. We also used these tools to search for relevant literature. No AI-generated text or suggestions were accepted without verification and editing by the authors. We did not use generative AI tools for research ideation or execution, drafting sections of the paper, generating synthetic datasets or proving mathematical claims. We take responsibility for the final content of this work, including any text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

Michelle R. Arkin and James A. Wells. Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nature Reviews Drug Discovery*, 3:301–317, 2004. doi: 10.1038/nrd1343.

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17981–17993, 2021.

Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *International Conference on Learning Representations*, 2024.

Marie-Jeanne Basse, Stéphane Betzi, Xavier Morelli, and Philippe Roche. 2P2Idb v2: update of a structural database dedicated to orthosteric modulation of protein–protein interactions. *Database*, 2016:baw007, 2016. doi: 10.1093/database/baw007.

Cédric Bouysset and Sébastien Fiorucci. Prolif: a library to encode molecular interactions as fingerprints. *Journal of Cheminformatics*, 13:72, 2021. doi: 10.1186/s13321-021-00548-6.

Yuyuan Chen, Shiyi Wang, Peter Potaptchik, Jaeyeon Kim, and Michael S. Albergo. Discrete tilt matching. In *International Conference on Machine Learning*, 2026.

Carles Corbi-Verge, Michael Garton, Satra Nim, and Philip M. Kim. Strategies to develop inhibitors of motif-mediated protein–protein interactions as drug leads. *Annual Review of Pharmacology and Toxicology*, 57:39–60, 2017. doi: 10.1146/annurev-pharmtox-010716-104805.

Julian Cremer, Ross Irwin, Alessandro Tibo, Jon Paul Janet, Simon Olsson, and Djork-Arné Clevert. FLOWR: Flow matching for structure-aware de novo, interaction- and fragment-based ligand generation. *Nature Computational Science*, 6(6):565–574, 2026. doi: 10.1038/s43588-026-00998-8.

Miruna Cretu, Charles Harris, Ilia Igashov, Arne Schneuing, Marwin Segler, Bruno Correia, Julien Roy, Emmanuel Bengio, and Pietro Liò. Synflownet: Design of diverse and novel molecules with synthesis constraints. In *International Conference on Learning Representations*, 2025.

Jérémy Desaphy, Eric Raimbaud, Pierre Ducrot, and Didier Rognan. Encoding protein–ligand interaction patterns in fingerprints and graphs. *Journal of Chemical Information and Modeling*, 53 (3):623–637, 2013. doi: 10.1021/ci300566n.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8780–8794, 2021.

Hongkun Dou, Zike Chen, Fengji Li, Hongjue Li, and Yue Deng. Plug-and-play guidance for discrete diffusion models via gradient-informed logit correction. In *International Conference on Machine Learning*, 2026.

Jerome Eberhardt, Diogo Santos-Martins, Andreas F. Tillack, and Stefano Forli. AutoDock Vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8):3891–3898, 2021. doi: 10.1021/acs.jcim.1c00203.

Paul G. Francoeur, Tomohide Masuda, Jocelyn Sunseri, Andrew Jia, Richard B. Iovanisci, Ian Snyder, and David R. Koes. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of Chemical Information and Modeling*, 60(9): 4200–4215, 2020. doi: 10.1021/acs.jcim.0c00411.

Jack F. Greenblatt, Bruce M. Alberts, and Nevan J. Krogan. Discovery and significance of protein–protein interactions in health and disease. *Cell*, 187(23):6501–6517, 2024. doi: 10.1016/j.cell.2024.10.038.

Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3d equivariant diffusion for target-aware molecule generation and affinity prediction. In *International Conference on Learning Representations*, 2023.

Xinheng He, Yijia Zhang, Haowei Lin, Xingang Peng, Xiangzhe Kong, Mingyu Li, and Jianzhu Ma. Peptide2mol: A diffusion model for generating small molecules as peptide mimics for targeted protein binding. *arXiv preprint arXiv:2511.04984*, 2025. doi: 10.48550/arXiv.2511.04984.

Zhilin Huang, Ling Yang, Xiangxin Zhou, Chujun Qin, Yijie Yu, Xiawu Zheng, Zikun Zhou, Wentao Zhang, Yu Wang, and Wenming Yang. Interaction-based retrieval-augmented diffusion models for protein-specific 3d molecule generation. In *International Conference on Machine Learning*, pp. 20348–20364, 2024.

Kazuyoshi Ikeda, Yuta Maezawa, Tomoki Yonezawa, Yugo Shimizu, Toshiyuki Tashiro, Satoru Kanai, Nobuyoshi Sugaya, Yoshiaki Masuda, Naoko Inoue, Tatsuya Niimi, Keiichi Masuya, Kenji Mizuguchi, Toshio Furuya, and Masanori Osawa. DLiP-PPI library: An integrated chemical database of small-to-medium-sized molecules targeting protein–protein interactions. *Frontiers in Chemistry*, 10:1090643, 2023. doi: 10.3389/fchem.2022.1090643.

Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In *International Conference on Learning Representations*, 2017.

Lingyan Jin, Weiru Wang, and Guowei Fang. Targeting protein–protein interaction by small molecules. *Annual Review of Pharmacology and Toxicology*, 54:435–456, 2014. doi: 10.1146/annurev-pharmtox-011613-140028.

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

Takatsugu Kosugi and Masahito Ohue. Quantitative estimate index for early-stage screening of compounds targeting protein–protein interactions. *International Journal of Molecular Sciences*, 22(20):10925, 2021. doi: 10.3390/ijms222010925.

Alston Lo, Connor W. Coley, and Wojciech Matusik. A genetic algorithm for navigating synthesizable molecular spaces. In *International Conference on Learning Representations*, 2026.

Hossam Nada, Yongseok Choi, Sungdo Kim, Kwon Su Jeong, Nicholas A. Meanwell, and Kyeong Lee. New insights into protein–protein interaction modulators in drug discovery and therapeutic advance. *Signal Transduction and Targeted Therapy*, 9:341, 2024. doi: 10.1038/s41392-024-02036-3.

Marta Pelay-Gimeno, Adrian Glas, Oliver Koch, and Tom N. Grossmann. Structure-based design of inhibitors of protein–protein interactions: Mimicking peptide binding epitopes. *Angewandte Chemie International Edition*, 54(31):8896–8927, 2015. doi: 10.1002/anie.201412070.

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2mol: Efficient molecular sampling based on 3d protein pockets. In *International Conference on Machine Learning*, volume 162, pp. 17644–17655, 2022.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, pp. 3942–3951, 2018. doi: 10.1609/aaai.v32i1.11671.

Yanru Qu, Keyue Qiu, Yuxuan Song, Jingjing Gong, Jiawei Han, Mingyue Zheng, Hao Zhou, and Wei-Ying Ma. MolCRAFT: Structure-based drug design in continuous parameter space. In *International Conference on Machine Learning*, volume 235, pp. 41749–41768, 2024.

Nir Qvit, Samuel J. S. Rubin, Travis J. Urban, Daria Mochly-Rosen, and Eric R. Gross. Peptidomimetic therapeutics: scientific approaches and opportunities. *Drug Discovery Today*, 22(2): 454–462, 2017. doi: 10.1016/j.drudis.2016.11.003.

Andrei Rekesh, Miruna Cretu, Dmytro Shevchuk, Vignesh Ram Somnath, Pietro Liò, Robert A. Batey, Mike Tyers, Michał Koziarski, and Cheng-Hao Liu. SynCoGen: Synthesizable 3d molecule generation via joint reaction and coordinate modeling. In *International Conference on Learning Representations*, 2026.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010. doi: 10.1021/ci100050t.

Seonghwan Seo, Minsu Kim, Tony Shen, Martin Ester, Jinkyoo Park, Sungsoo Ahn, and Woo Youn Kim. Generative flows on synthetic pathway for drug design. In *International Conference on Learning Representations*, 2025.

Tony Shen, Seonghwan Seo, Ross Irwin, Kieran Didi, Simon Olsson, Woo Youn Kim, and Martin Ester. Compositional flows for 3d molecule and synthesis pathway co-design. *arXiv preprint arXiv:2504.08051*, 2025. doi: 10.48550/arXiv.2504.08051.

Heqi Sun, Jiayi Li, Yufang Zhang, Shenggeng Lin, Junwei Chen, Hong Tan, Ruixuan Wang, Xueying Mao, Jianwei Zhao, Rongpei Li, and Dong-Qing Wei. Hot-spot-guided generative deep learning for drug-like ppi inhibitor design. *Interdisciplinary Sciences: Computational Life Sciences*, 18:180–194, 2026. doi: 10.1007/s12539-025-00756-w.

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

Andy R. D. Voet and Kai Zhang. Pharmacophore modelling as a virtual screening tool for the discovery of small molecule protein–protein interaction inhibitors. *Current Pharmaceutical Design*, 18:4586–4598, 2012. doi: 10.2174/138161212802651616.

Chenyu Wang, Masatoshi Uehara, Yichun He, Amy Wang, Avantika Lal, Tommi Jaakkola, Sergey Levine, Aviv Regev, Hanchen Wang, and Tommaso Biancalani. Fine-tuning discrete diffusion models via reward optimization with applications to dna and protein design. In *International Conference on Learning Representations*, 2025.

James A. Wells and Christopher L. McClendon. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature*, 450:1001–1009, 2007. doi: 10.1038/nature06526.

Zeyu Wen, Jiahua He, Huanyu Tao, and Sheng-You Huang. PepBDB: a comprehensive structural database of biological peptide–protein interactions. *Bioinformatics*, 35(1):175–177, 2019. doi: 10.1093/bioinformatics/bty579.

Rongting Yue, Zekun Yang, Gustavo Seabra, Chenglong Li, and Yanjun Li. Pep2mol: 3d molecule generation targeting protein–protein interfaces with diffusion models. *bioRxiv*, 2026. doi: 10.64898/2026.06.28.734975.

Wonho Zhung, Hyeongwoo Kim, and Woo Youn Kim. 3d molecular generative framework for interaction-guided drug design. *Nature Communications*, 15:2688, 2024. doi: 10.1038/s41467-024-47011-2.

Appendix

A Training Details

A.1 Interface Encoder

We provide a schematic of the target-conditioned cross-modal interface encoder in Figure 5.

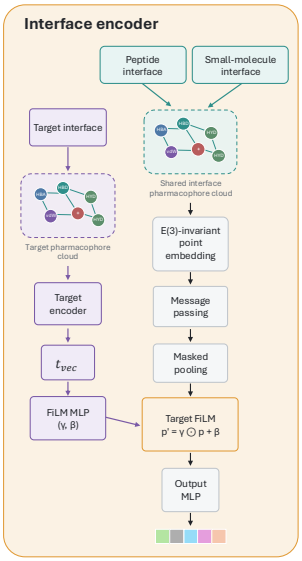


Figure 5: **Target-conditioned Cross-Modal Interface encoder architecture.**

A.1.1 Dataset Construction

We apply the following filtering, deduplication, partitioning, and query-construction procedures to the four structural resources described in Section 3.2.1.

Candidates are merged across sources and filtered by: (i) at least four target-contact residues; (ii) for small molecules, at least four heavy atoms and one ProLIF-detected interaction type; (iii) for the lower-confidence source, a Vina score at most -6.0, at least four eligible poses per interface, and retention of the top 25% by score; (iv) an interface point cloud (Section 3.2.2) containing at least four points in the shared frame; and (v) cross-source deduplication, retaining the higher-confidence instance when sources agree. We additionally cap cloud size at the train-split 99.5th percentile: 237 points for peptides and 37 for small molecules.

We partition data at the UniProt-accession level. Accessions are grouped into connected components under pairwise GRIM pharmacophore similarity (Desaphy et al., 2013) at threshold 0.80, then assigned to a fixed-seed $80/20$ train/test split; no accession occurs in both partitions. The resulting corpus contains 3,090 peptide instances from 271 accessions (2,578 train; 512 test) and 97,669 small-molecule instances from 292 accessions (73,494 train; 24,175 test). Restricting representation training to examples with extractable target clouds gives 2,467 peptide and 72,919 small-molecule instances across 215 shared accessions.

For target-conditioned alignment, a peptide query is eligible when it has a real target cloud and at least four same-target positive and four negative molecular endpoints. Labels are computed from contact profiles in the same curated pool: a positive endpoint has nonzero target-contact and interaction-type agreement with the query; a negative has zero agreement and a pocket-centroid distance below 15 Å in the shared frame. This produces 82 eligible queries across 18 targets. Held-out evaluation uses the disjoint accession-level test partition, further restricted to pairs with extractable target clouds and usable interaction profiles.

A.1.2 HYPERPARAMETERS

Here we summarize the default hyperparameters used for training the interface encoder.

Table 4: **Default hyperparameters.**

Hyperparameters	Values
Interface encoder hidden dimension	128
Interface embedding dimension	64
Target encoder	hidden dim. 64, output dim. 32
Stage I: modality-shared geometric pretraining	
Epochs	30
Steps per epoch	2,000
Batch size	128
Learning rate	1×10^{-4}
Weight decay	1×10^{-4}
Partial-view margin m_p	1.0
Stage II: target-conditioned cross-modal alignment	
Training cycles	150
Queries per rank	2
Learning rate	3×10^{-5}
Min. positive / hard examples	4 / 4
Examples per label	4
Max. queries per target	6
Ranking margin	0.1
Temperature	0.1
γ-regularization weight	0.2
Target-vector diversity weight	1.0
Target-vector diversity margin	0.3

A.2 PPI-ORIENTED PRIOR ADAPTATION

For target-specific prior adaptation, let $V(x)$ be AutoDock Vina score, $Q(x)$ QEPPI score (Kosugi & Ohue, 2021), and $h(x)$ heavy-atom count. Positive reward used in implementation is

$$R_{\mathrm{PPI}}(x) = \max(-0.1V(x), 10^{-5})Q(x)S_{\mathrm{size}}(h(x)), \tag{17}$$

$$S_{\mathrm{size}}(h) = \begin{cases} 1, & h \le 45, \\ \exp[-0.15(h-45)], & h > 45. \end{cases} \tag{18}$$

DTM uses log reward $r(x) = \log \max(R_{\mathrm{PPI}}(x), 10^{-12})$, centered by current replay-pool mean before tilting. This yields target distribution $p_R(x) \propto p_0(x) \exp(\beta r(x))$.

For masked position i, let $\pi_a(\cdot \mid x_t, i)$ be frozen reference posterior from previous tilt phase, π_θ trainable posterior, x_1^i completed rollout token, h local tilt increment, and c distillation coefficient. Reward-induced local target is

$$T_c(v \mid x_t, i, x_1) = \frac{c\pi_a(v \mid x_t, i) + [\exp(hr(x_1)) - c]\mathbf{1}[v = x_1^i]}{\exp(hr(x_1))}. \tag{19}$$

Corresponding weighted cross-entropy can be evaluated without normalizing sequence-level tilted likelihood:

$$\exp(hr)\,\mathrm{CE}(T_c, \pi_\theta) = c\,\mathrm{CE}(\pi_a, \pi_\theta) - [\exp(hr) - c]\log \pi_\theta(x_1^i \mid x_t, i). \tag{20}$$

Same objective is applied to masked node and edge decisions alongside original coordinate-flow objective. This adaptation changes generic synthesis-aware prior into target-specific PPI prior but uses no peptide reference.

A.3 JOINT GUIDANCE WITH PEPTIDE REFERENCE

A.3.1 HYPERPARAMETERS

Here we provide the hyperparameters used for the PIFT sampling procedure.

Table 5: **Default hyperparameters.**

Hyperparameters	Discrete branch	Coordinate branch
Guidance start t_{start}	0.5	0.5
Schedule exponent p	2.0	2.0
Gradient clipping	5.0	5.0
Guidance scale $\gamma_{\max}$	0.1	0.1
Margin fraction	1.2	—
ST-Gumbel MC worlds M	4	—
ST-Gumbel temperature τ	1.0	—

A.3.2 Gradient Stabilization

The magnitude and direction of the interface-loss gradient can vary substantially across reverse-sampling steps as the molecular state becomes progressively resolved. Applying these raw gradients directly can therefore produce abrupt corrections that perturb the pretrained sampling trajectory. We stabilize the guidance signal using RMS normalization and clipping to control its instantaneous scale, followed by Adam-style moment smoothing to reduce step-to-step variation.

For a preprocessed gradient $\tilde{\mathbf{g}}_t$, we maintain

$$\begin{aligned} \mathbf{m}_t &= \beta_1 \mathbf{m}_{t-1} + (1-\beta_1)\tilde{\mathbf{g}}_t, \\ \mathbf{v}_t &= \beta_2 \mathbf{v}_{t-1} + (1-\beta_2)\tilde{\mathbf{g}}_t^{\odot 2}, \\ \hat{\mathbf{g}}_t &= \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t} + \epsilon}, \end{aligned} \tag{21}$$

where $\hat{\mathbf{m}}_t$ and $\hat{\mathbf{v}}_t$ are bias-corrected moments. The stabilized direction $\hat{\mathbf{g}}_t$ is then used in the corresponding logit or coordinate update.

B EXPERIMENTAL DETAILS

B.1 BASELINES

Below, we summarize the core generative mechanisms and conditioning strategies of the baseline methods used in our evaluation.

For pocket-conditioned 3D generation, **TargetDiff** (Guan et al., 2023) uses $\mathrm{SE}(3)$-equivariant diffusion to jointly generate ligand atom types and 3D coordinates conditioned on protein-pocket atoms, whereas **MolCRAFT** (Qu et al., 2024) models molecular geometry and discrete chemical attributes in a unified continuous parameter space using a pocket-conditioned Bayesian Flow Network. **FLOWR** (Cremer et al., 2026) combines continuous coordinate flow matching with categorical atom and bond flows and integrates a separately encoded protein pocket through equivariant cross-attention; it can additionally condition on desired protein–ligand interaction features.

For synthesis-aware generation, pocket-specific **CGFlow** (Shen et al., 2025) jointly searches over reaction-based molecular construction and 3D binding poses within the target complex using compositional generative flows that combine a GFlowNet for molecular construction with flow matching for coordinates, while **SynCoGen** (Rekesh et al., 2026) couples absorbing-state discrete diffusion over graph-structured building-block and reaction representations with continuous flow matching over coordinates. **SynGA** (Lo et al., 2026) instead performs target-specific genetic optimization directly over synthesis routes using synthesizability-preserving crossover and mutation operators.

For PPI-specific conditioning, **Hot2Mol** (Sun et al., 2026) integrates a conditional transformer with a variational autoencoder to generate molecules, whereas **Peptide2Mol** (He et al., 2025) conditions an $\mathrm{E}(3)$-equivariant diffusion model jointly on a peptide binder and its surrounding protein environment to generate peptide-informed 3D small molecules.

B.2 HYPERPARAMETERS

Here we summarize the default hyperparameters used for sampling experiments.

Table 6: **Default hyperparameters.**

Hyperparameters	Values
Reverse sampling steps	100
Batch size	64
Samples per target	256
Docking	UniDock
Docking seeds	5

C Extended Results

C.1 IFA Validation

We validate Interface Agreement (IFA) against the overlap of active target residues on held-out peptide–molecule pairs. Residue-set Jaccard similarity correlates with mean IFA with pooled Pearson $r = 0.987$ and query-wise Spearman $\rho = 0.963$. The compatibility-restricted IFA variant shows similarly strong agreement ($r = 0.996$, $\rho = 0.985$), supporting IFA as a graded measure of peptide–molecule interaction recovery.

C.2 Interface Encoder Architecture Ablation

We examine how target conditioning FiLM layer affects cross-modal organization in the learned interface space. We compare the target-agnostic encoder with concatenation-based conditioning, Early FiLM, and the Late-FiLM design used in our final model. For each held-out peptide query, we evaluate whether a compatible small-molecule interface is embedded closer than a mismatched interface and measure the corresponding positive–negative distance margin. As shown in Table 7, Late FiLM yields the highest pair-ranking accuracy and the largest pair margin among the tested variants.

Table 7: **Ablation of target conditioning.** Pair ranking measures the fraction of peptide queries for which the compatible small-molecule interface is embedded closer than its mismatched counterpart. Pair margin is the difference between mismatched- and compatible-pair distances, $d_{\text{neg}} - d_{\text{pos}}$, where larger values indicate stronger separation.

Encoder variant	Pair ranking ↑	Pair margin ↑
No target conditioning	0.572 (± 0.029)	0.067 (± 0.032)
Concatenation	0.586 (± 0.030)	0.054 (± 0.056)
Early FiLM	0.579 (± 0.018)	0.044 (± 0.024)
PIFT (Late FiLM)	**0.598** (± 0.042)	**0.099** (± 0.064)

C.3 Reference Specificity Analysis

We study reference-specific guidance behavior on MDM2 using two experimentally resolved peptide binders, 1YCR and 3LNJ, whose target-contact patterns partially overlap but remain distinguishable. Figure 6 summarizes reference-specific effects at the score, structural, and population levels. Each guided population is evaluated against both its own reference and the alternate peptide. Relative to the guidance-off population, 1YCR-guided sampling shows a larger IFA improvement when evaluated against 1YCR than against 3LNJ, while 3LNJ-guided sampling exhibits the converse pattern. This differential response indicates that guidance is sensitive to the selected peptide interaction pattern rather than merely promoting a generic MDM2-compatible interface. Representative binding modes provide a structural view of this effect: molecules generated from the two references occupy similar regions of MDM2 but adopt distinct local interaction geometries around the peptide-contacting surface. At the population level, differential residue-by-interaction fingerprints over the observed hydrophobic and van der Waals interaction channels reveal localized shifts between the two guided ensembles, further supporting reference-dependent modulation of the target interface.

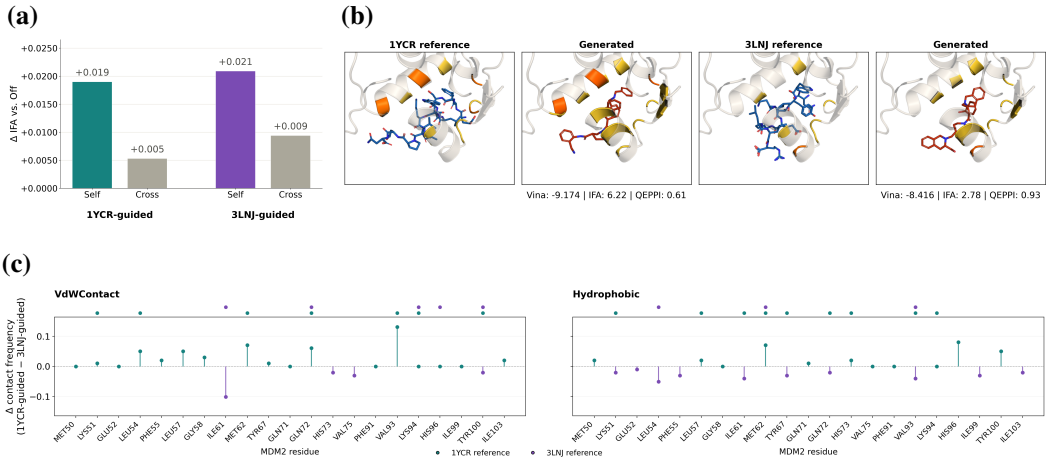


Figure 6: **Reference-specificity of peptide guidance on MDM2.** (a) Self- and cross-reference IFA scoring. Changes in IFA relative to Off for the 1YCR- and 3LNJ-guided ensembles, evaluated against their own (self) or alternate (cross) peptide reference. Self-reference gains are larger for both guided ensembles. (b) Reference peptides and representative guided binding modes. Guided molecules occupy similar MDM2 regions while adopting reference-dependent local interaction geometries. (c) Differential population-level interaction fingerprints. Localized shifts reveal reference-dependent interaction patterns.

C.4 Full Results for PPI targets

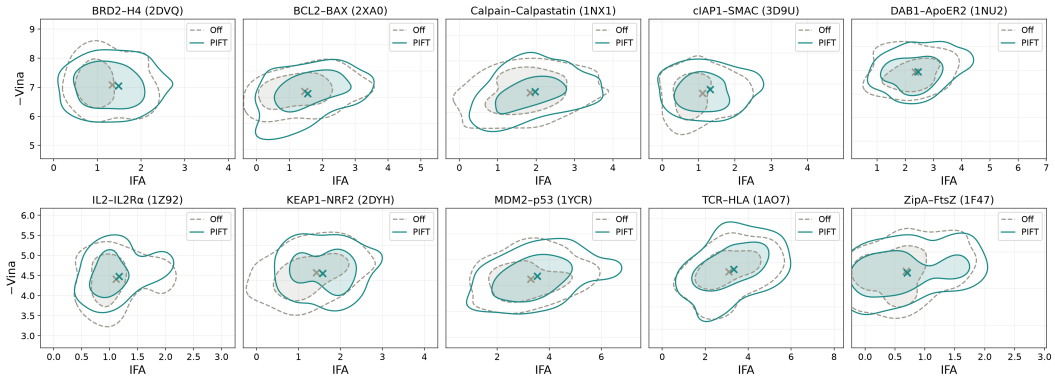


Figure 7: **Full target-level results for guided and unguided molecular populations in IFA–−Vina space.**

Table 8: **Average Vina docking score for 10 PPI targets.** Best results are in bold.

Category	Model	Average Vina Docking Score (kcal/mol, ↓) BRD2–H4	BCL2–BAX	Calpain–Calpastatin	cIAP1–SMAC	DAB1–ApoER2
Autoregressive	Hot2Mol	-7.00 (± 0.04)	-7.61 (± 0.05)	-6.56 (± 0.03)	-6.59 (± 0.05)	-7.35 (± 0.02)
Search-based	SynGA	-6.47 (± 0.09)	-7.33 (± 0.05)	-6.50 (± 0.03)	-6.04 (± 0.09)	-6.78 (± 0.03)
GFlowNet-based	CGFlow	**-7.20** (± 0.05)	-7.93 (± 0.03)	-6.91 (± 0.03)	-6.38 (± 0.03)	-6.99 (± 0.05)
Diffusion / Flow-based	TargetDiff	-5.30 (± 0.04)	–	-6.74 (± 0.03)	-3.50 (± 0.03)	-6.02 (± 0.08)
	MolCRAFT	-5.02 (± 0.02)	-6.84 (± 0.03)	-7.50 (± 0.04)	-3.90 (± 0.04)	-5.32 (± 0.01)
	Flowr	-4.00 (± 0.19)	-1.25 (± 0.00)†	-5.00 (± 0.08)	-6.39 (± 0.05)	-6.04 (± 0.21)
	SynCoGen	-6.68 (± 0.08)	-7.70 (± 0.04)	-7.00 (± 0.05)	-6.33 (± 0.02)	-7.05 (± 0.08)
	Peptide2Mol	-5.68 (± 0.54)	–	-6.18 (± 0.33)	-5.12 (± 0.23)	-4.90 (± 0.07)
	PIFT	-6.99 (± 0.04)	**-8.22** (± 0.06)	**-7.75** (± 0.06)	**-6.78** (± 0.05)	**-7.92** (± 0.04)

Category	Model	Average Vina Docking Score (kcal/mol, ↓) IL2–IL2R	KEAP1–NRF2	MDM2–p53	TCR–HLA	ZipA–FtsZ
Autoregressive	Hot2Mol	-4.35 (± 0.04)	-7.12 (± 0.05)	-7.54 (± 0.03)	-9.30 (± 0.01)	-6.69 (± 0.03)
Search-based	SynGA	-4.04 (± 0.05)	-6.93 (± 0.04)	-6.94 (± 0.07)	-8.77 (± 0.06)	-6.37 (± 0.04)
GFlowNet-based	CGFlow	-4.26 (± 0.04)	-7.48 (± 0.05)	-7.53 (± 0.04)	-8.86 (± 0.03)	-6.75 (± 0.04)
Diffusion / Flow-based	TargetDiff	–	-6.90 (± 0.04)	-6.95 (± 0.06)	-7.83 (± 0.06)	-6.94 (± 0.07)
	MolCRAFT	-3.30 (± 0.04)	-6.80 (± 0.05)	-6.96 (± 0.06)	-6.50 (± 0.10)	-5.99 (± 0.03)
	Flowr	**-4.61** (± 0.03)	-7.13 (± 0.02)	-7.93 (± 0.05)	-9.24 (± 0.03)	-5.77 (± 0.03)
	SynCoGen	-3.98 (± 0.05)	-7.46 (± 0.05)	-6.99 (± 0.04)	-9.33 (± 0.04)	-6.63 (± 0.06)
	Peptide2Mol	–	-5.59 (± 0.07)	-5.60 (± 0.08)	–	-5.47 (± 0.04)
	PIFT	-4.42 (± 0.05)	**-8.26** (± 0.06)	**-8.40** (± 0.06)	**-10.65** (± 0.06)	**-7.37** (± 0.05)

Docking efficiency. To ensure that docking improvements are not solely driven by larger molecular size, we regress Vina score against heavy-atom count within each target and compare mean residuals across methods (Figure 8). PIFT remains favorable after size correction across most targets and achieves a strongly negative mean residual overall, indicating that its docking advantage does not rely on increasing molecular-size. Combined with the increased IFA, this supports selective transfer of peptide-defined interaction patterns rather than size-based mimicry.

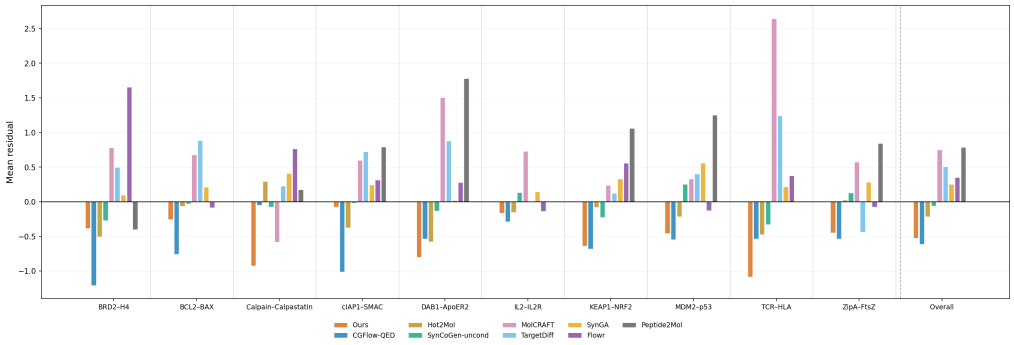


Figure 8: **Heavy-atom-count-adjusted docking comparison.** Bars show mean residual Vina score after fitting a separate Vina–heavy-atom-count regression within each target; the final group averages residuals across all 10 targets. Lower residuals indicate more favorable docking scores than expected at matched molecular size.

Table 9: **PoseCheck metrics.** Average interaction counts for the 10-target benchmark, restricted to models with successful evaluation on all 10 targets. Values are means ± standard deviations over three runs; the main-text table reports only the corresponding Sum column. Best results are in bold, and the second-best H-bond acceptor value is underlined.

Model	PoseCheck Metrics (↑)				
	H-Bond Acceptors	H-Bond Donors	Van der Waals	Hydrophobic	Sum
Hot2Mol	0.201 (± 0.004)	0.128 (± 0.008)	7.187 (± 0.044)	2.968 (± 0.046)	10.485 (± 0.066)
SynGA	0.241 (± 0.024)	0.071 (± 0.006)	7.797 (± 0.051)	2.961 (± 0.045)	11.071 (± 0.088)
CGFlow	0.130 (± 0.006)	0.054 (± 0.009)	6.583 (± 0.032)	2.427 (± 0.013)	9.193 (± 0.048)
MolCRAFT	0.232 (± 0.022)	**0.141** (± 0.024)	7.736 (± 0.093)	3.231 (± 0.044)	11.340 (± 0.124)
SynCoGen	**0.262** (± 0.012)	0.139 (± 0.008)	7.364 (± 0.041)	2.883 (± 0.037)	10.648 (± 0.071)
PIFT	<u>0.251</u> (± 0.018)	0.109 (± 0.007)	**8.115** (± 0.037)	**3.875** (± 0.040)	**12.351** (± 0.066)