SpecLig: Energy-Guided Hierarchical Model for Target-Specific 3D Ligand Design

Peidong Zhang[1,2], **Rong Han**[1,2], **Xiangzhe Kong**[1, 4], **Ting Chen**[1,2,*], **and Jianzhu Ma**[3,4,*]

[1]Department of Computer Science and Technology, Tsinghua University, Beijing, China
[2]Institute of Artificial Intelligence, Tsinghua University, Beijing, China
[3]Department of Electronic Engineering, Tsinghua University, Beijing, China
[4]Institute for AI Industry Research, Tsinghua University, Beijing, China
[*]Correspondence should be addressed to tingchen@mail.tsinghua.edu.cn and majianzhu@tsinghua.edu.cn

ABSTRACT

Structure-based generative models often optimize single-target affinity with ignorance of specificity, resulting in the generation of high-affinity candidates that exhibit promiscuous binding across unrelated targets. This decoupling of affinity and specificity not only compromises therapeutic efficacy but also elevates off-target risks that constrain translational potential. Therefore, we introduce SpecLig, a unified structure-based framework that jointly generates small molecules and peptides with improved target affinity and specificity. SpecLig represents a complex as a block-based graph, combining a hierarchical SE(3)-equivariant variational autoencoder with an energy-guided geometric latent-diffusion model. Chemical priors derived from block–block contact statistics are explicitly incorporated, biasing generation towards pocket-complementary fragment combinations. We benchmark SpecLig on peptide and small-molecule tasks using standard public datasets and propose precision/breadth testing paradigms to quantify specificity. Across multiple evaluations, ligand candidates generated by SpecLig usually bind to the target pocket with high specificity and affinity while maintaining competitive advantages in other attributes. Ablations indicate that both hierarchical representation and energy guidance contribute to success. Finally, we present multiple real applications that demonstrate how SpecLig improves ligands in natural complexes to mitigate potential off-target risks. SpecLig, therefore, provides a practical route to prioritize higher-specificity designs for downstream experimental validation. The codes are available at: https://github.com/CQ-zhang-2016/SpecLig.

1 Introduction

Designing ligands with target specificity advances both mechanistic biology and drug discovery[1,2]. Recent deep-learning generative models[3–8] have shifted ligand discovery from large-scale screening[9–12] toward rational, structure-guided design. Structure-based drug design (SBDD) is central to this trend because it leverages the receptor's three-dimensional (3D) geometry to constrain generation in chemically and spatially relevant regions of ligand space[13–15].

Ligands have diverse chemical modalities, including small molecules, peptides, and antibodies. Each class has distinct pharmacokinetic and functional profiles (e.g., oral bioavailability for small molecules[16]; recognition motifs for peptides[17]). Despite differences in scale and physicochemical properties, ligand–receptor binding follows the same physicochemical rules, such as bond types, bond angles, lengths, and steric clashes[18,19]. Common generative paradigms such as autoregressive models[20–22], diffusion-based methods[23–26], flow-matching[27–29], retrieval-augmented generation[30,31], energy optimization[32–35], iterative refinement[36–39], and voxel-based models[40,41] have been successfully applied across these ligand classes. Recent works[42–46] have considered unified generative frameworks that accommodate representational differences among ligand types.

Off-target is a well known challenge in traditional drug discovery[47–50]. Most SBDD models learn ligand distributions conditioned on a single target structure and, therefore, emphasize recurring motifs from the training data. While such motifs can increase predicted affinity, their systematic reuse can reduce target specificity. For examples (Figure 1a and b), voxbind and PepGLAD produced candidates with higher predicted affinity than native ligands for receptors (UniProt IDs: P14779 and P33247); however, they also showed enhanced binding to unrelated proteins. This decoupling of affinity and specificity indicates a systematic bias toward promiscuous binders, increasing off-target risk and constraining translational potential[51–53]. In the interaction analyses shown in Figure 1a and 1b, some fragments in the designed ligands contributed little to binding with the intended target yet dominated interactions with off-targets. We further analyzed several advanced SBDD models and categorized their outputs by specificity. As shown in Figure 1c, small molecules with low specificity tend to contain a markedly higher fraction (≈5–10%) of polar groups, which promote promiscuous binding across multiple targets. Figure 1d shows that highly specific peptides typically contain a higher fraction (≈3–10%) of helical structure, whereas non-helical flexible segments increase exposure risk. Experimental details are provided in Supplementary Chapter 3. These pervasive generative biases in deep learning models highlight the need to explicitly consider specificity in ligand design.

Standard SBDD benchmarks focus on single-target docking scores, measuring spatial and electrostatic complementarity[54,55]. Generative models can produce candidates with high docking scores through excessive modifications, masking potential promiscuity. Prior attempts[56] to quantify specificity in small-molecule design by comparing against a random non-target are limited by randomness and provide little actionable guidance. To address this, we design a suite of specificity metrics that quantify binding preferences across multi-target contexts and are applicable to both small-molecule and peptide design. Using these metrics, we find that many model-generated ligands with improved single-target affinity exhibit reduced specificity relative to reference ligands. It is consistent with the empirical difficulty of improving selectivity versus affinity[57].

We think that achieving target specificity requires moving beyond single-structure conditioning to incorporate evolutionary binding preferences. We address this by introducing energy-guided diffusion that leverages statistical contact frequencies between molecular fragments derived from native protein-ligand complexes. Unlike conventional energy functions based on physical potentials, our potential terms quantify the empirical preference for specific inter-fragment interactions across diverse targets. Based on this, we propose SpecLig (shown in Figure 1), a unified generative framework that integrates structure-based design with statistical energy guidance to simultaneously design high-specificity small molecules and peptides. SpecLig employs a hierarchical graph neural network (GNN): an atom-level encoder for local chemistry and bond order, and a block-level encoder that represents residues or predefined molecular fragments to capture global topology with reduced cost. Fragments are derived by subgraph partitioning[58] of a molecular library to form a reusable vocabulary. Crucially, chemical priors derived from native complexes are incorporated both as features in training and as additive guidance during latent diffusion sampling. It biases generation toward a chemical space that is more likely to enhance specific binding under the given conditions, improving binding affinity while preserving geometric accuracy.

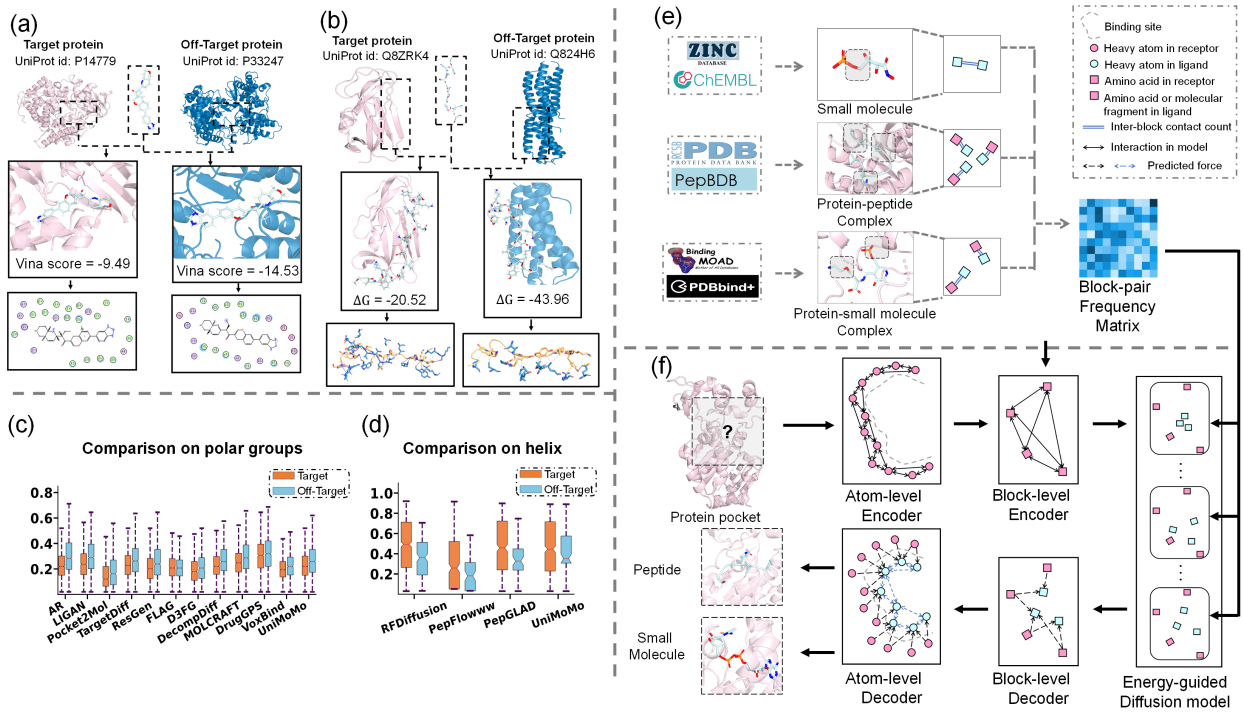

Figure 1. Off-target analyses and the main architecture of SpecLig. We present two representative cases in (a) and (b). Existing SBDD models generate small-molecule and peptide ligands with high predicted affinity for their intended receptors (pink) but also exhibit undesired binding to unrelated proteins (blue). (c) and (d) provide a comparative analysis of chemical and structural features distinguishing high- and low-specificity designs. (c) differences in polar-group ratios among small molecules, and (d) differences in helical content among peptides. (e) and (f) show the overview of the proposed SpecLig framework. (e) depicts the construction of the block–block frequency matrix used as a statistical energy prior, and (f) illustrates how this prior guides the latent diffusion–based generation process within the hierarchical VAE architecture of SpecLig.

We evaluated SpecLig on both peptide and small molecule benchmarks. The results show that our unified framework attains competitive performance on standard benchmarks while consistently producing ligands with higher specificity. Moreover, we demonstrate that SpecLig can simultaneously optimize affinity and specificity for native ligands that exhibit potential off-target risks, illustrating its practical utility in advancing candidate molecules toward safer leads.

2 Methods

SpecLig addresses the critical challenge of binding specificity in structure-based ligand design through statistical energy guidance. SpecLig models protein–ligand complexes as graphs of molecular fragments (blocks) and combines a hierarchical variational autoencoder (VAE[59]) with an E(3)-equivariant latent diffusion model. A hierarchical encoder extracts multi-scale geometric representations, and a decoder with a similar architecture reconstructs block types and full-atom 3D coordinates. The core innovation lies in an energy-guided diffusion process that leverages empirical block–block interaction statistics from natural complexes, steering generation toward target-specific binding configurations while maintaining high affinity. Supplementary Chapter 2 discusses related work on SpecLig.

2.1 Notations and overview of SpecLig

We denote a protein–ligand complex as a block-based graph $G = (V, E)$. Each node, $v_i \in V$, is an unordered set of atoms represented as $v_i = \left\{ (\alpha_i, \vec{x}_i) \,\middle|\, \alpha_i \in \mathbb{Z}^{n_i},\ \vec{x}_i \in \mathbb{R}^{n_i \times 3} \right\}$, consisting of all the element types and

3D coordinates for the n_i atoms in the ith block. Edges E record intra- and inter-block chemical bonds and spatial adjacency. A block vocabulary S (canonical amino-acid residues and predefined small-molecule fragments) assigns each block a type $s_i \in S$. For controllable generation, we introduce a flag $p_i \in \{0, 1\}$ in each block: $p_i = 1$ restricts sampling to canonical residues; otherwise, it samples from the full vocabulary. The pre-computed empirical block–block frequency matrix is denoted as $\mathbf{F} \in \mathbb{R}^{n_S \times n_S}$, where n_s is the length of S.

For conditional design, the pocket and the ligand are represented as G_P and G_L, and the task is to model the conditional generation distribution $p(G_L \mid G_P)$. Assuming that ξ, ϕ, and θ respectively represent the parameter sets of the VAE encoder, decoder, and latent diffusion model, the generation probability can be expressed in marginalized form:

$$p(G_L \mid G_P) = \int p_\theta\big(Z_L^0 \mid Z_P\big)\; p_\phi\big(G_L \mid Z_L^0, Z_P\big)\; dZ_L^0, \tag{1}$$

$$p_\theta\big(Z_L^0 \mid Z_P\big) = \int p\big(Z_L^T\big) \prod_{t=T}^{1} p_\theta\big(Z_L^{t-1} \mid Z_L^t, Z_P\big)\; dZ_L^{1:T}, \tag{2}$$

In Equations 1 and 2, Z_P denotes the latent representation of the pocket, Z_L^0 denotes the noise-free latent representation of the ligand, and Z_L^T is the terminal latent state approaching the prior (Gaussian noise). Finally, the pocket is defined as blocks whose reference points (C_β or the fragment centroid) lie within 10Å of any native ligand atom, preventing information leakage[21].

2.2 Implementation of hierarchical VAE

We decompose the VAE into hierarchical modules $\mathcal{E}_\xi = \{\, \mathcal{E}_{\xi,1},\ \mathcal{E}_{\xi,2} \,\}$ and $\mathcal{D}_\phi = \{\, \mathcal{D}_{\phi,1},\ \mathcal{D}_{\phi,2} \,\}$. All networks are implemented with SE(3)-equivariant transformers[60] to guarantee translation–rotation equivariance.

Atom-scale to block-scale encoding. The atom-scale encoder $E_{\xi,1}$ encodes each atom with its element type, parent block type, canonical-residue flag, and chain id. We augment atom features with a correlated projection learned from the empirical frequency matrix $\mathbf{F}$ after temperature-scaled normalization.

As shown in Figure 1f, the information flow is restricted to operate within G_L and G_P separately when constructing the k-nearest neighbors (KNN) graph. Edge features are defined as $\{\, e_{ab},\ \mathbf{d}_{ab},\ \boldsymbol{\beta}_{ab} \,\}$, indicating whether the two atoms belong to the same block, their relative distance, and any candidate chemical-bond types. The block-scale encoder $E_{\xi,2}$ constructs a coarser KNN graph on top of $E_{\xi,1}$'s outputs. The edge features are updated to the relative distances between block centroids. Define $\{\, \boldsymbol{\mu}_i,\ \boldsymbol{\sigma}_i,\ \vec{\boldsymbol{\mu}}_i,\ \vec{\boldsymbol{\sigma}}_i \,\}$ as the distributional parameters learned by the encoders. After reparameterized sampling[61], we could obtain the block-level latent representations for G_L and G_P, expressed as:

$$Z_P = \mathcal{E}_{\xi,2}(\mathcal{E}_{\xi,1}(G_P))\,, \quad Z_L^0 = \mathcal{E}_{\xi,2}(\mathcal{E}_{\xi,1}(G_L))\,, \tag{3}$$

The latent representation of ith block could be defined as $\big\{(\mathbf{z}_i, \vec{\mathbf{z}}_i) \big|\, \mathbf{z}_i \in \mathbb{R}^d,\ \vec{\mathbf{z}}_i \in \mathbb{R}^3 \,\big\}$, where d is set to 8. Before decoding, Gaussian noise is injected to model the deviations encountered during the latent diffusion process, thereby improving robustness.

Block-scale to atom-scale decoding. The hierarchical decoders decode block types and reconstruct atomic coordinates at their respective scales. Information flow is constrained to be unidirectional from G_P to G_L (shown in Figure 1f). The block-scale decoder predicts block type probabilities and coarse block centroid positions. Outputs are normalized with softmax. Querying the block vocabulary S to retrieve the set of atom types, intra-block bonds, and the initial atomic positions. Specifically, design task is identified based on user input, and the block type prediction is explicitly restricted to the relevant vocabulary.

$$h,\ \vec{r} = \mathcal{D}_{\phi,2}(Z_P, Z_L)\,, \quad s = \mathrm{MLP}_{\mathrm{block}}(h), \tag{4}$$

We then adopt an iterative, flow-matching–like decoding scheme[62]. For each block i and each time step t, we predict two atom-level features h_i^t and v_i^t, which indicate per-atom feature representations and relative coordinate updates, respectively. Initial atom positions $\vec{\mathbf{x}}_i^1$ are sampled randomly around the predicted centroids. Time step t decreases from 1.0 to 0.0 with $\Delta t = 0.1$. The update dynamics follow:

$$h_i^t,\ \vec{v}_i^t = \mathcal{D}_{\phi,1}\Bigg(\Big\{\boldsymbol{\alpha}_i,\ \boldsymbol{\beta}_i,\ \vec{\mathbf{x}}_i^t\Big\},\ h_i\Bigg), \tag{5}$$

$$\vec{\mathbf{x}}_i^{\,t-\Delta t} = \vec{\mathbf{x}}_i^{\,t} + \Delta t\, \vec{v}_i^{\,t}, \tag{6}$$

Chemical bond prediction. From the final atomic representations $\{h_i^0, \vec{\mathbf{x}}_i^0\}$ produced by the hierarchical decoder, we predict inter-block bond distributions using a two-stage procedure (shown in Equations 7 and 8). A small MLP-based frontier predictor first selects spatially-proximal (distance < 3.5Å), inter-block atom pairs. Then another MLP-based bond predictor is used.

$$p(a,b) = \delta\left(\mathrm{MLP}_{\mathrm{focal}}\left(h_{i,a}^0, h_{j,b}^0 \mid d\left(\vec{\mathbf{x}}_i^0, \vec{\mathbf{x}}_j^0\right)\right)\right), \tag{7}$$

$$\boldsymbol{\beta}_{ij,\,ab} = \sigma\left(\mathrm{MLP}_{\mathrm{bond}}\left(h_{i,a}^0,\, h_{j,b}^0 \mid p(a,b)\right)\right), \tag{8}$$

Where (a,b) are the indices of two different atoms in the ith block and the jth block, separately. $d(\cdot)$, $\delta(\cdot)$, and $\sigma(\cdot)$ respectively indicate the distance calculation, gated function, and softmax activation. In practice, predicted bond assignments are accepted based on confidence ranking and valence compatibility. To refine geometric and chemical accuracy, the entire encoder–decoder pipeline may be iteratively run for multiple design cycles, depending on the task.

2.3 Energy-guided geometric diffusion model

We perform conditional diffusion in the latent space and inject a data-driven energy guidance during sampling that explicitly favors fragment pairings empirically associated with cognate binding. A vanilla SpecLig variant without energy guidance is reported in Supplementary Figure S1 as an ablation.

Statistical prior construction. As illustrated in Figure 1e, we precomputed empirical block–block contact priors from multiple curated sources. (i) Fragment-fragment occurrences from over a million small molecules sampled from ZINC15[63] and ChEMBL[64] databases, decomposed by the principal-subgraph algorithm[58]. (ii) Residue pairs bound by inter-chain hydrogen bonding from RCSB PDB[65] and PepBDB[66] databases, using trajectory-based hydrogen-bond analysis software. (iii) protein–ligand interaction frequencies from PDBbind[67] and Binding-MOAD[68] analyzed with BINANA. Source-wise frequency matrices are normalized to reduce modal bias and log-transformed to yield continuous statistical potentials. Noted that the data used to construct the frequency matrix were strictly filtered to entries published prior to 2020, ensuring no overlap with the evaluation benchmarks. Rather than memorizing specific complex structures, it serves as a coarse-grained statistical prior, conceptually analogous to BLOSUM[69] matrices in protein sequence analysis. The frequency matrix captures universal, target-independent bonding probabilities between amino acids and molecular fragments, leveraging recurring physicochemical patterns to guide generation.

Why does it improve specificity. Frequent fragments in training data often reflect general potential rather than pocket specificity. Using $\mathbf{F}$ as a statistical potential reweights sampling toward fragment combinations that historically co-occur in native complexes for pockets similar to the query, thereby reducing the generation of promiscuity-prone motifs.

Standard diffusion process. We adopt a standard diffusion forward–backward scheme. Let the encoder outputs for the pocket and ligand (Z_P and Z_L) be the clean latent variables Z_P^0 and Z_L^0, respectively. Z_L^0 is normalized following empirical practices[70]. We apply a discrete-time Gaussian forward noising process with a cosine schedule $\{\beta_t\}_{t=1}^T$, then define $\alpha_t = 1 - \beta_t$ and $\overline{\alpha}_t = \prod_{k=1}^t \alpha_k$. Using the attribute latents z_i^t as an example, the forward Markov chain is given by:

$$q\left(\mathbf{z}_i^t \,\middle|\, \mathbf{z}_i^{t-1}\right) = \mathcal{N}\left(\mathbf{z}_i^t;\, \sqrt{\alpha_t}\,\mathbf{z}_i^{t-1},\, \beta_t \mathbf{I}\right) \tag{9}$$

$$q\left(\mathbf{z}_i^t \,\middle|\, \mathbf{z}_i^0\right) = \mathcal{N}\left(\mathbf{z}_i^t;\, \sqrt{\overline{\alpha}_t}\,\mathbf{z}_i^0,\, (1-\overline{\alpha}_t)\,\mathbf{I}\right) \tag{10}$$

The reverse denoising is parameterized conditionally as:

$$p_\theta\left(\mathbf{z}_i^{t-1} \,\middle|\, Z_L^t, Z_P^0\right) = \mathcal{N}\left(\mathbf{z}_i^{t-1};\, \mathbf{z}_\theta\left(Z_L^t, Z_P^0, t\right),\, \beta_t \mathbf{I}\right) \tag{11}$$

$$\mathbf{z}_\theta\big(Z_L^t, Z_P^0, t\big) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{z}_i^t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\,\boldsymbol{\varepsilon}_\theta\big(Z_L^t, Z_P^0, t\big)\,[i]\right) \tag{12}$$

Where $\boldsymbol{\varepsilon}_\theta$ also uses the SE(3)-equivariant network. The latent representation of block position $\vec{\boldsymbol{z}}_i^{\,t}$ follows the same forward–backward scheme.

Energy integration into sampling. During sampling, we guide each reverse step with an energy term derived from the block–block bonding prior to favoring chemically plausible ligands. Given a denoised

estimate $\hat{Z}_L^0$ reconstructed from the current predicted noise, we decode block-type probability vectors $\hat{s}$ with the frozen $\mathcal{D}_{\phi,2}$ following Equation 4. For a block pair (i, j), the pairwise energy term is defined as:

$$E_{ij} = -\log\left(\left(\hat{\mathbf{s}}_i^\top \mathbf{F} \hat{\mathbf{s}}_j\right)^{1/\tau}\right) \cdot \omega_{ij} \tag{13}$$

Where τ is a temperature smoothing factor, and ω_{ij} is a distance-dependent decay. For stability, we apply thresholding to the predicted probabilities. The total energy E is normalized based on the ligand's molecular mass. Crucially, this energy term quantifies how well the current fragment combination matches the target-selective binding patterns observed in nature.

We integrate the energy into the reverse dynamics by backpropagating the energy gradient from the block probabilities into the noise space. From the reverse format of Equation 10, we could obtain the denoised representation $\hat{Z}_L^0$ reconstructed by the current predicted noise $\left(\hat{\boldsymbol{\varepsilon}}_t, \hat{\vec{\boldsymbol{\varepsilon}}}_t\right)$. Thus, the backpropagation follows:

$$g_\varepsilon = \sum_{i<j} \nabla_{\left(\hat{\boldsymbol{\varepsilon}}_t,\, \hat{\vec{\boldsymbol{\varepsilon}}}_t\right)} E = \sum_{i<j} \nabla_{\hat{\mathbf{s}}} E \cdot \nabla_{\hat{Z}_L^0} \hat{\mathbf{s}} \cdot \nabla_{\left(\hat{\boldsymbol{\varepsilon}}_t,\, \hat{\vec{\boldsymbol{\varepsilon}}}_t\right)} \hat{Z}_L^0, \tag{14}$$

Gradient update only the noise-prediction network. Using the attribute latents as an example, we adjust the sampling process by stepping in the negative energy gradient direction:

$$\hat{\boldsymbol{\varepsilon}}_t = \hat{\boldsymbol{\varepsilon}}_t - \omega_s \cdot \omega_t \cdot g_\varepsilon \cdot \min\left(1,\, \frac{G_{\text{set}}}{\|g_\varepsilon\|_2}\right) \tag{15}$$

To guarantee stability, gradients are clipped by their norm to a preset bound G_{set}, and the guidance weight ω_t decays with time steps. The weight ω_s is explicitly assigned based on the task type to modulate the overall response intensity. Details of the sensitivity analysis for ω_s can be found in Supplementary Chapter 13. Besides, pseudocodes for the full training and guided sampling algorithms are provided in Supplementary Chapters 4 and 5.

2.4 Training procedure and loss design

We design a composite loss to optimize multiple objectives. The hierarchical VAE and the latent diffusion model are trained sequentially, following best practices[42,43]. In the hierarchical VAE stage, atom-scale loss terms include: (i) a focal atom-pair classification loss and bond-type cross-entropy for inter-block bond prediction, (ii) a mean-squared error (MSE) loss on the predicted velocity fields produced during iterative decoding, and (iii) a paired-distance loss between atom pairs. The velocity-field supervision is obtained by interpolating between noisy and ground-truth coordinates. The paired-distance loss is applied only for adjacent (≤ 6Å) atoms in the native structure and during early decoding steps ($t \leq 0.25$).

At the block scale, we impose Kullback–Leibler (KL) regularization on both attribute latents and coordinate latents, and supervise block-type classification and coarse centroid regression. Lastly, we consider the contrastive loss to align the ligand and target pockets from a global perspective. Global descriptors are obtained by averaging per-node features. The triplet-based contrastive loss pulls the ligand descriptors closer to their corresponding pocket and pushes them away from a randomly sampled pocket. The training objective is the weighted sum of the atom-scale, block-scale, and global contrastive terms. The training uses teacher forcing: atomic types, intra-block bonds, and 50% of inter-block bonds are exposed to the VAE. Additionally, 5% of pocket residues are also masked during training.

For the latent diffusion stage, we train with a weighted sum of the denoising MSE loss and a latent perceptual loss (LPL). The LPL term follows the original formulation[71], and the detailed implementation is provided in Supplementary Chapter 6. The overall diffusion-stage loss follows below. Lastly, we provide the default hyperparameters in Supplementary Chapter 9:

$$\mathcal{L}_{\text{LDM}} = \mathbb{E}_{t\sim \text{U}(1,\ldots,T)}\left[\|\boldsymbol{\varepsilon}_t - \hat{\boldsymbol{\varepsilon}}_t\|_2^2 + \left\|\vec{\boldsymbol{\varepsilon}}_t - \hat{\vec{\boldsymbol{\varepsilon}}}_t\right\|_2^2\right] + \mathcal{L}_{\text{lpl}}\left(\hat{Z}_L^0,\, Z_L^0\right) \tag{16}$$

3 Experiments

We trained and evaluated SpecLig on peptide (PepBench[42], ProtFrag[42], and LNR[72]) and small-molecule (CrossDocked2020[73]) design benchmarks using standardized protocols, following the splitting recommendations

provided by the authors[20,42]. Considering the splitting bias suggested by work[74], we provided the similarity between the training and test sets in Supplementary Chapter 1. Domain-specific baselines were trained only on their respective domain data, whereas the unified frameworks were trained on the multi-domain training data. To enhance small-molecule chemistry modeling, we augmented training with 100,000 ChEMBL[64] compounds under 20% structural masking. All dataset details are provided in Supplementary Chapter 1.

Evaluation employed commonly used domain metrics grouped by evaluation objective. We report (i) per-category weighted ranking scores in overall comparisons and (ii) per-metric numerical results within each category. Ranking scores were computed as $(N - \overline{r})$, where N is the number of competing models in each category and $\overline{r}$ is the mean rank across the category's individual metrics. Notably, the weights across metric categories in our overall comparisons were manually assigned to prioritize target specificity within the broader context of this multi-objective optimization task. All numerical results reported represent per-sample means calculated across the respective test sets. They were normalized to $[0, 1]$ for clarity, the higher, the better.

We introduce two paradigms to quantify specificity. The precision paradigm docks designs against the target and one randomly chosen non-target using a high-accuracy engine; the breadth paradigm screens designs against a protein set that includes the target to mimic multi-target pharmacological environments. For peptides, we use PyRosetta side-chain docking (multiple precision modes, set size 100); for small molecules, we use Uni-Dock[75] (breadth scans, set size 2000) and AutoDock Vina[76] (precision). Finally, comprehensive ablations appear in Supplementary Chapter 10.

3.1 Small molecules

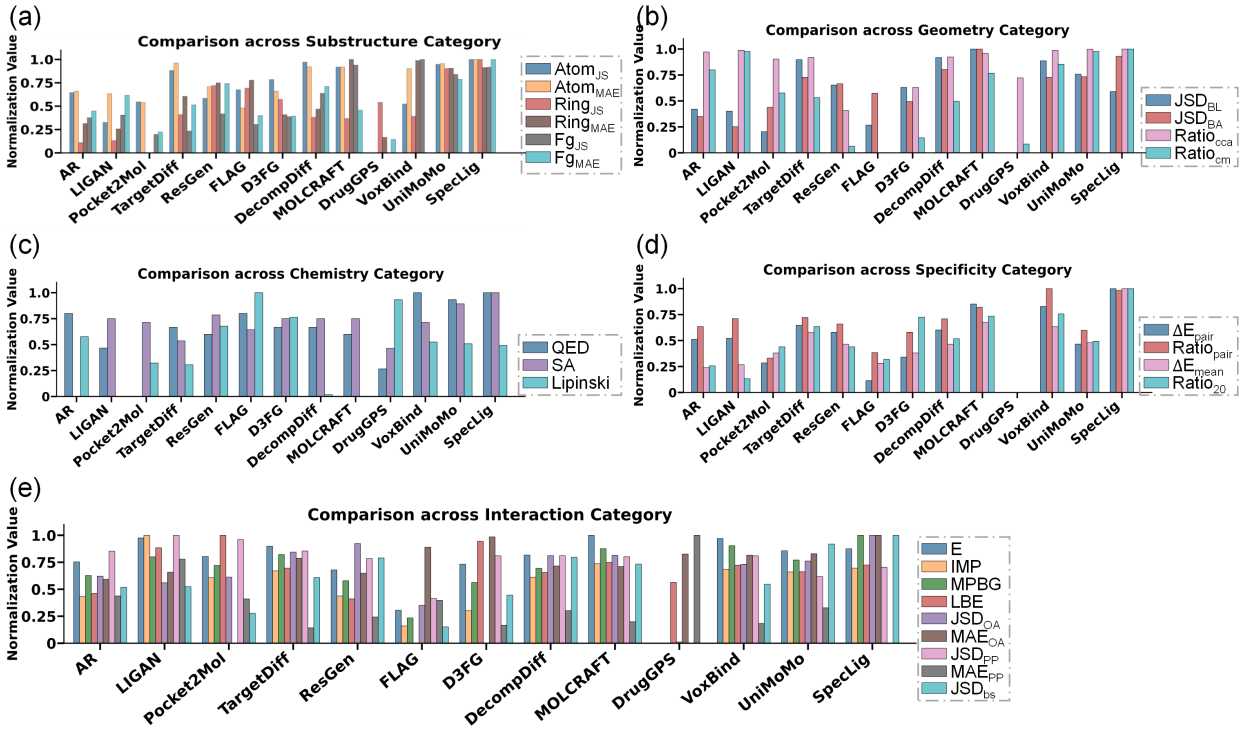

Figure 2. Performance comparison of small-molecule generative models across five categories of metrics: Substructure (a), Geometry (b), Chemistry (c), Specificity (d), and Interaction (e).

Setup. We follow prior benchmarks[77] and group metrics into five categories: **Substructure**, **Geometry**, **Chemistry**, **Specificity**, and **Interaction**. **Substructure** measures agreement at the atom-type, ring, and functional-group levels using the Jensen–Shannon divergence (JSD, ↓) for distributional shift and per-molecule mean absolute error (MAE, ↓) for frequency deviations. **Geometry** assesses local geometric validity via the JSD of bond lengths (JSD_{BL}, ↓) and bond angles (JSD_{BA}, ↓), as well as atom-collision statistics: mean collided-atom fraction ($\text{Ratio}_{\text{cca}}$, ↓) and the fraction of collided molecules (Ratio_{cm}, ↓). **Chemistry** quantifies

drug-likeness (QED, ↑), normalized synthetic accessibility (SA, ↑), LogP ($-0.4 \lesssim$ ideal $\lesssim 5.6$), and the ratio satisfying Lipinski rules (LPSK, ↑). In **Specificity**, $\Delta\mathbf{E}_{\text{pair}}$ (↓) and $\text{Ratio}_{\text{pair}}$ (↑) report the score difference between target and non-target dockings and the fraction of cases in which the target scores better, respectively; $\Delta\mathbf{E}_{\text{mean}}$ (↓) and Ratio_{20} (↑) report the extent to which the target score exceeds the set mean and the ratio of designs in the top 20%. **Interaction** evaluates binding performance: mean AutoDock Vina score (E, ↓) and fraction exceeding the reference ligand (IMP, ↑). To correct for the effects of molecular size, we report the per-pocket mean relative energy gain (MPBG, ↑) and ligand binding efficiency normalized per atom (LBE, ↑). Interaction patterns are further characterized via seven PLIP-detected[78] interaction types: population-level agreement is quantified by JSD_{OA} (↓) and MAE_{OA} (↓), per-pocket agreement by JSD_{PP} (↓) and MAE_{PP} (↓). Besides, JSD_{bs} (↓) compares the fragment composition in the binding site.

Baselines. We compare SpecLig to representative methods with various generative paradigms: 3D autoregressive (AR[20], Pocket2Mol[21], ResGen[22]), diffusion-based (TargetDiff[24], DecompDiff[25]), fragment-based (FLAG[36], D3FG[37], DrugGPS[39]), voxel-based (LiGAN[40], VoxBind[41]), and continuous-space approaches (MolCRAFT[79], UniMoMo[43]). Baseline designs follow the CBGBench and MOLCRAFT benchmarks where available; others use standard reported settings.

Table 1. Overall comparisons for de novo small molecule design.

Model	Substruct. 0.15	Chem. 0.15	Interact. 0.25	Geom. 0.15	Speci. 0.30	**Rank**
AR	0.56	0.86	1.18	0.83	0.96	10
LIGAN	0.56	0.60	<u>2.08</u>	0.90	1.26	8
Pocket2Mol	0.20	0.45	1.50	0.53	0.96	11
TargetDiff	1.00	0.50	1.83	1.05	2.61	5
ResGen	1.10	1.16	1.15	0.56	1.80	7
FLAG	0.78	1.16	0.70	0.26	0.66	12
D3FG	0.80	1.20	1.15	0.56	1.41	9
DecompDiff	1.16	0.75	1.63	1.16	2.10	6
MOLCRAFT	1.23	0.50	2.00	1.43	3.15	3
DrugGPS	0.23	0.65	0.68	0.18	0.00	13
VoxBind	0.93	1.10	1.80	1.28	<u>3.21</u>	<u>2</u>
UniMoMo	<u>1.50</u>	<u>1.35</u>	1.73	<u>1.46</u>	1.65	4
SpecLig	**1.70**	**1.46**	**2.15**	**1.50**	**3.51**	**1**

Results and discussion. Table 1 and Figure 2 summarize weighted scores and per-metric ranks (complete results in Supplementary Chapter 7). SpecLig exhibits robust gains across multiple categories. On Specificity, it ranks first or second across metrics ($\Delta\mathbf{E}_{\text{pair}} = -0.83$, $\text{Ratio}_{\text{pair}} = 58.73\%$, $\Delta\mathbf{E}_{\text{mean}} = -0.75$, $\text{Ratio}_{20} = 30.17\%$, Supplementary Table S3), indicating preferential binding to intended pockets. In Interaction metrics, SpecLig achieves near-optimal mean Vina scores and the highest MPBG ($= 15.17$) , representing a 53.4% relative gain versus VoxBind (Supplementary Table S5). Chemistry and Substructure metrics remain competitive, implying improved pocket specificity without loss of drug-likeness or synthetic accessibility. Geometrically, collision rates are low, though bond-length distributions still show room for refinement.

Despite these advances, gains in specificity and affinity are markedly smaller for small molecules than for peptides. This reflects fundamental chemical and sampling differences. The Small-molecule space is discrete and higher-dimensional: combinatorial variation in functional groups, ring systems, and rotatable bonds yields a highly multimodal energy landscape that a single block-level vocabulary sparsely covers. By contrast, amino-acid semantics are compact, so frequency-derived priors transfer more effectively. Small-molecule binding is also far more sensitive to substructure details. Therefore, overly aggressive empirical guidance induces mode collapse, while weak guidance leaves sampling inertia and reduces specificity. Thus, block–block frequency priors are a useful foundation but are insufficient alone. For small molecules, we therefore advocate integrating richer, physics-informed cues: explicit force-field or electrostatic terms, conformational-ensemble priors, fragment-aware flexible docking, adaptive guidance schedules, and targeted local geometry refinement.

We believe that integrating multi-objective scoring can compensate for the limitations of frequency priors in capturing geometric details, thereby achieving a better balance between affinity and specificity.

3.2 Peptides

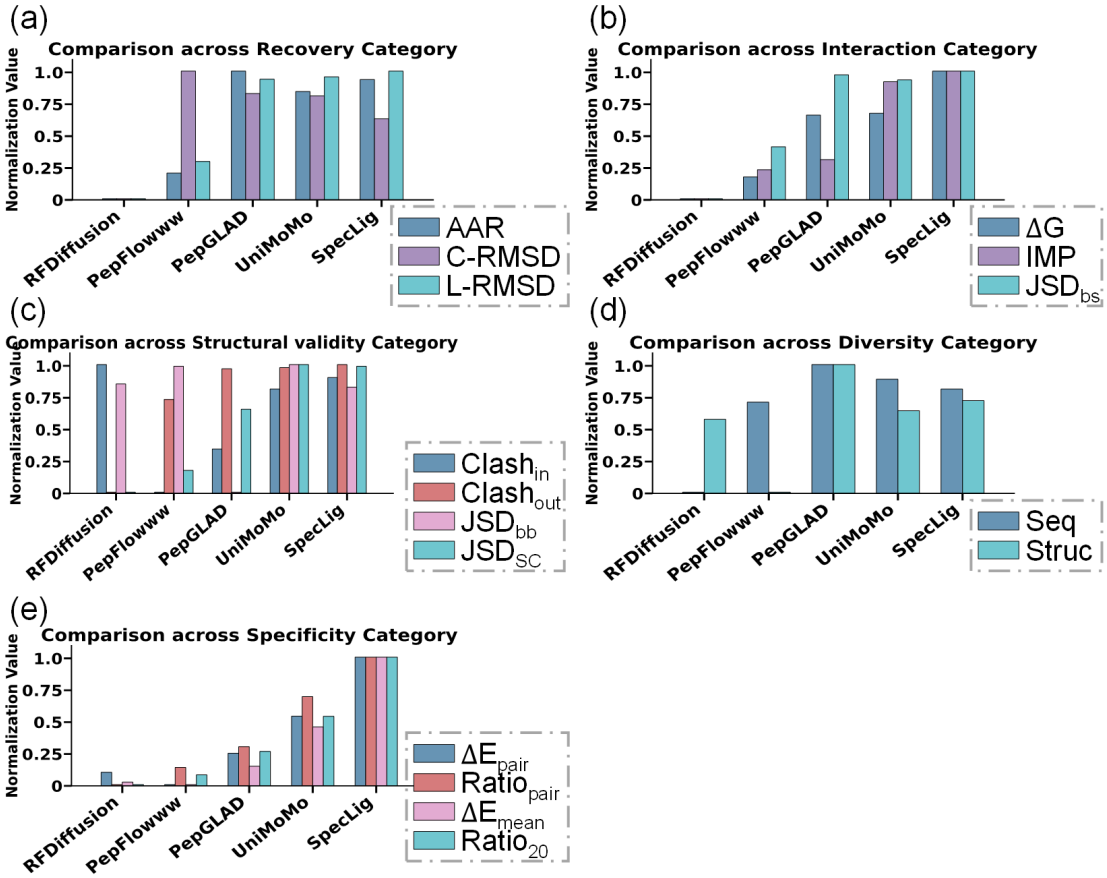

Figure 3. Comparison of peptide design models across five metric categories: Recovery (a), Interaction (b), Structural validity (c), Diversity (d), and Specificity (e).

Setup. Referring to benchmark[43], we organize metrics into five categories: **Specificity**, **Recovery**, **Interaction**, **Structural Validity**, and **Diversity**. **Specificity** follows the protocol described above. **Recovery** measures agree with the reference sequence and conformation. Amino-acid recovery (AAR, ↑) is the per-position sequence match rate. conformational recovery includes complex-aligned RMSD (C-RMSD, ↓) and ligand-aligned RMSD (L-RMSD, ↓). **Interaction** is primarily evaluated by the mean binding energy computed with PyRosetta (ΔG, ↓) and by the proportion of designed peptides that outperform the native ligand on their targets (IMP, ↑). We also report JSD_{bs}, ↓ following the same definition. **Structural validity** is assessed via atomic clash rates: Clash_{in} (↓) and $\text{Clash}_{\text{out}}$ (↓) record intra-peptide and peptide–receptor residue collisions (collision defined as C_α distance < 3.6574Å). Besides, backbone and side-chain dihedral deviations are summarized by JSD_{bb} (↓) and JSD_{SC} (↓) using 10° bins. **Diversity** is reported as the ratio of unique clusters to total designs (cluster threshold: sequence identity $> 40\%$ and RMSD < 2Å).

Baselines.

Peptide baselines include RFDiffusion[23], PepFlow[27], PepGLAD[42], and UniMoMo[43]. Our RFDiffusion follows the published two-stage pipeline but is limited to a single inverse-fold and relax cycle to avoid extra optimization that could bias physical metrics. PepFlow implements multimodal flow-matching for joint sequence–structure generation. PepGLAD performs latent diffusion to produce sequences and all-atom conformations. UniMoMo is the unified multimodal baseline. All methods were executed according to the recommended settings.

Table 2. Overall comparisons for de novo peptide design.

Model	Recov. 0.15	Struc. 0.15	Interact. 0.25	Div. 0.15	Speci. 0.30	**Rank**
RFDiffusion	0.00	0.23	0.00	0.08	0.15	5
PepFlowww	0.30	0.18	0.25	0.08	0.15	4
PepGLAD	**0.45**	0.18	0.60	**0.60**	0.60	3
UniMoMo	0.35	**0.48**	0.68	0.38	0.90	2
SpecLig	0.41	0.41	**1.00**	0.38	**1.20**	**1**

Results and discussion. Peptides, as larger ligands, are handled consistently by SpecLig. Table 2 and Figure 3 summarize SpecLig's performance on the peptide task, and the complete results are shown in Chapter 8 of the Supplementary Information. SpecLig shows a clear advantage in the peptide task, especially in Specificity and Interaction. In Specificity (Supplementary Table S8), SpecLig increases the second-best model's Ratiopair and Ratio20 from 68.75% and 52.91% to 75.43% and 75.00%, respectively; native ligands score 80.72% and 78.31%. In Interaction benchmarks (Supplementary Table S11), SpecLig is the only method with a negative mean $\Delta G = -1.92$, compared to 29.21 for the runner-up UniMomo. These gains reflect reduced off-target binding and higher-affinity designs. SpecLig also attains the lowest $\text{Clash}_{\text{out}}$ and L-RMSD, indicating geometrically self-consistent outputs. A modest diversity reduction is observed, which we attribute to the constraining effect of energy guidance.

Mechanistically, the gains stem from two design choices. Firstly, the hierarchical representation preserves fragment semantics and filters atom-level noise, simplifying the identification and localization of critical local units. Secondly, incorporating interaction priors into latent diffusion biases sampling toward chemically plausible, pocket-complementary solutions. Ablation studies shown in Supplementary Figure S1 confirm that removing either component degrades Specificity and Interaction. In short-peptide tasks, we observe a clear diversity–specificity tradeoff (Supplementary Figure S5): increasing energy guidance steadily reduces sequence diversity, while affinity and specificity initially rise but fall at high guidance strengths. The late decline is consistent with sampling-induced mode collapse, where over-constraining gradients concentrate probability mass on few modes and suppress beneficial exploration. In this work, we empirically tuned the guidance parameters to near-optimal values for the two design tasks. Nonetheless, SpecLig stands to benefit from more advanced schemes, such as adaptive guidance schedules implemented via gradient callbacks, entropy-aware regularization, or homology-guided template retrieval, to better reconcile specificity with diversity. Overall, SpecLig strikes a favorable balance across competing objectives, prioritizing specificity and affinity, thereby materially improving downstream candidate triage.

3.3 Case study

Two case studies demonstrate SpecLig's reduction of off-target risk in small-molecule (Figure 4a–d) and peptide (Figure 4e–h) design. In both cases, SpecLig designs yielded no valid docking poses on non-target proteins (Figure 4d and 4h).

For Cytochrome P450BM-3 (UniProt: P14779), the native ligand exhibits moderate affinity (Vina=-6.08; Figure 4a), driven by a carboxylate-mediated hydrogen-bond network with residues 47, 72, 73 and hydrophobic alkyl chain contacts. However, it binds more strongly to the non-target aldehyde decarbonylase (UniProt: Q54764; Vina=-7.32; Figure 4b) due to its topology and reusable carboxylate motif. In contrast, SpecLig's design incorporates a thiadiazole fragment that forms directional hydrogen bonds with residues 330 and 437, meanwhile enhancing hydrophobic contacts (Vina=-9.58; Figure 4c). No candidate pose satisfied the required geometric constraints in the non-target pocket (Figure 4d), confirming SpecLig's optimization of pocket-specific geometric and chemical complementarity.

Against the Ferrichrome-iron receptor (UniProt: P06971), native microcin J25 forms four target hydrogen bonds (Figure 4e) but nine with non-target rhodopsin (UniProt: P02699; Figure 4f), resulting in a stronger off-target affinity (-30.46) than target (-8.71). It illustrates that promiscuous contact motifs may lead to off-target risks. The SpecLig-designed peptide forms up to twelve hydrogen bonds in the target pocket, accompanied by extensive hydrophobic contacts (Figure 4g, $\Delta G = -60.21$), and fails to adopt a feasible

pose on the off-target (the binding energy was higher than that of the respective monomeric states, shown in Figure 4h). It demonstrates the simultaneous improvement of target affinity and suppression of non-specific binding.

Methodologically, coupling block-level modeling with an empirical energy prior shifts the model from learning generic interfacial fragments to performing pocket-conditioned fragment selection, thereby producing chemically and geometrically customized ligands. Importantly, SpecLig does not merely fill pockets by agglomerating fragments. Analyses in Supplementary Chapter 11 indicate an adaptive sizing mechanism within SpecLig: it is pocket topology, rather than blocks' initial setting, that predominantly determines the designed ligand's size. Lastly, additional target examples are provided in Supplementary Chapter 12.

4 Conclusion

SpecLig demonstrates that integrating block-wise chemical priors with hierarchical equivariant modeling yields ligand designs that better balance affinity and specificity than existing generative baselines. The hierarchical VAE reduces atom-level noise and preserves fragment semantics, while energy-guided latent sampling steers generation toward pocket-conditioned, chemically plausible solutions. Unlike black-box models, SpecLig's preference for certain motifs can be cross-referenced with known biophysical principles, providing a rationale for the generated 3D structures. Empirical evaluation suggests consistent, interpretable improvements in specificity and interaction in both peptide and small-molecule design. Ablations verify the complementary roles of each module. Some limitations also remain. Gains in small-molecule design are constrained by discrete chemical complexity and geometric sensitivity, suggesting opportunities to incorporate richer physical cues (force-field terms, electrostatics, or conformational ensembles). Besides, prospective experimental validations are still in need. Overall, SpecLig is a pioneering step toward designing ligands with enhanced specificity and may inspire future advancements in this promising direction.

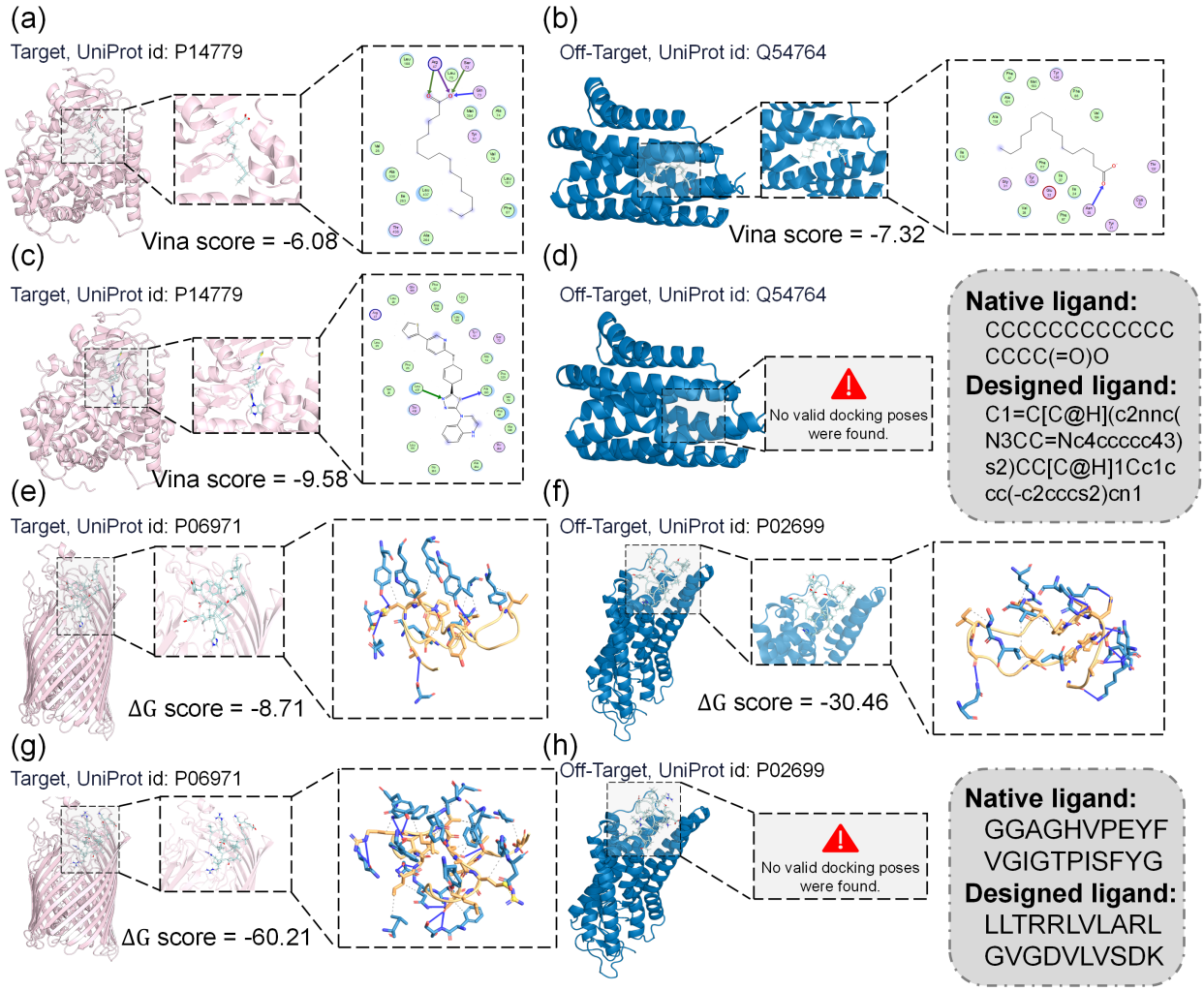

Figure 4. Case studies illustrating SpecLig's reduction of off-target binding. Native small molecule targeting cytochrome P450BM-3 (a, pink) and its non-target aldehyde decarbonylase (b, blue). (c) and (d) show SpecLig-designed small molecule binds with target and non-target, separately. Similarly, (e-f) and (g-h) show the native peptide and the designed peptide binding to the target ferrichrome-iron receptor (pink) and the non-target rhodopsin (blue), respectively. All docking scores were computed using Schrödinger Glide and Rosetta under high-precision settings. Interaction analyses were performed with MOE for small molecules and the newest PLIP for peptides. The sequences of native and designed ligands are given in gray boxes.

References

1. Glögl, M. *et al.* Target-conditioned diffusion generates potent tnfr superfamily antagonists and agonists. *Science* **386**, 1154–1161 (2024).
2. Wang, Z. *et al.* Naturally enhanced neutralizing breadth against sars-cov-2 one year after infection. *Nature* **595**, 426–431 (2021).
3. Jumper, J. *et al.* Highly accurate protein structure prediction with alphafold. *nature* **596**, 583–589 (2021).
4. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
5. Brandes, N., Ofer, D., Peleg, Y., Rappoport, N. & Linial, M. Proteinbert: a universal deep-learning model of protein sequence and function. *Bioinformatics* **38**, 2102–2110 (2022).
6. Ross, J. *et al.* Large-scale chemical language representations capture molecular structure and properties. *Nature Machine Intelligence* **4**, 1256–1264 (2022).
7. Zhang, P., Xia, C. & Shen, H.-B. High-accuracy protein model quality assessment using attention graph neural networks. *Briefings Bioinformatics* **24**, bbac614 (2023).
8. Gujral, O., Bafna, M., Alm, E. & Berger, B. Sparse autoencoders uncover biologically interpretable features in protein language model representations. *Proceedings National Academy Sciences* **122**, e2506316122 (2025).
9. Gao, B. *et al.* Drugclip: Contrasive protein-molecule representation learning for virtual screening. *Advances Neural Information Processing Systems* **36** (2024).
10. Singh, R., Sledzieski, S., Bryson, B., Cowen, L. & Berger, B. Contrastive learning in protein language space predicts interactions between drugs and protein targets. *Proceedings National Academy Sciences* **120**, e2220778120 (2023).
11. Zhang, P., Ma, J. & Chen, T. Escaping the drug-bias trap: using debiasing design to improve interpretability and generalization of drug-target interaction prediction. *IEEE Transactions on Computational Biology Bioinformatics* (2025).
12. Honarparvar, B., Govender, T., Maguire, G. E., Soliman, M. E. & Kruger, H. G. Integrated approach to structure-based enzymatic drug design: molecular modeling, spectroscopy, and experimental bioactivity. *Chemical reviews* **114**, 493–537 (2014).
13. Xie, W., Wang, F., Li, Y., Lai, L. & Pei, J. Advances and challenges in de novo drug design using three-dimensional deep generative models. *Journal Chemical Information Modeling* **62**, 2269–2279 (2022).
14. Du, Y. *et al.* Machine learning-aided generative molecular design. *Nature Machine Intelligence* **6**, 589–604 (2024).
15. Bepler, T. & Berger, B. Learning protein sequence embeddings using information from structure. *arXiv preprint arXiv:1902.08661* (2019).
16. Beck, H., Härter, M., Haß, B., Schmeck, C. & Baerfacker, L. Small molecules and their impact in drug discovery: A perspective on the occasion of the 125th anniversary of the bayer chemical research laboratory. *Drug Discovery Today* **27**, 1560–1574 (2022).
17. Derakhshankhah, H. & Jafari, S. Cell penetrating peptides: A concise review with emphasis on biomedical applications. *Biomedicine & Pharmacotherapy* **108**, 1090–1096 (2018).
18. Bissantz, C., Kuhn, B. & Stahl, M. A medicinal chemist's guide to molecular interactions. *Journal medicinal chemistry* **53**, 5061–5084 (2010).
19. Haynes, W. M. *CRC handbook of chemistry and physics* (CRC press, 2016).
20. Luo, S., Guan, J., Ma, J. & Peng, J. A 3d generative model for structure-based drug design. *Advances Neural Information Processing Systems* **34**, 6229–6239 (2021).
21. Peng, X. *et al.* Pocket2mol: Efficient molecular sampling based on 3d protein pockets. In *International Conference on Machine Learning*, 17644–17655 (PMLR, 2022).

22. Zhang, O. *et al.* Resgen is a pocket-aware 3d molecular generation model based on parallel multiscale modelling. *Nature Machine Intelligence* **5**, 1020–1030 (2023).

23. Watson, J. L. *et al.* De novo design of protein structure and function with rfdiffusion. *Nature* **620**, 1089–1100 (2023).

24. Guan, J. *et al.* 3d equivariant diffusion for target-aware molecule generation and affinity prediction. *arXiv preprint arXiv:2303.03543* (2023).

25. Guan, J. *et al.* Decompdiff: diffusion models with decomposed priors for structure-based drug design. *arXiv preprint arXiv:2403.07902* (2024).

26. Xie, X., Valiente, P. A., Kim, J. & Kim, P. M. Helixdiff, a score-based diffusion model for generating all-atom α-helical structures. *ACS Central Science* **10**, 1001–1011 (2024).

27. Lin, H. *et al.* Ppflow: Target-aware peptide design with torsional flow matching. *arXiv preprint arXiv:2405.06642* (2024).

28. Wu, F. *et al.* D-flow: Multi-modality flow matching for d-peptide design. *arXiv preprint arXiv:2411.10618* (2024).

29. Jing, B., Berger, B. & Jaakkola, T. Alphafold meets flow matching for generating protein ensembles. *arXiv preprint arXiv:2402.04845* (2024).

30. Zhang, P., Peng, X., Han, R., Chen, T. & Ma, J. Rag2mol: structure-based drug design based on retrieval augmented generation. *Briefings Bioinformatics* **26** (2025).

31. Lee, S. *et al.* Molecule generation with fragment retrieval augmentation. *Advances Neural Information Processing Systems* **37**, 132463–132490 (2024).

32. Gu, S. *et al.* Aligning target-aware molecule diffusion models with exact energy optimization. *Advances Neural Information Processing Systems* **37**, 44040–44063 (2024).

33. Huang, Z. *et al.* Protein-ligand interaction prior for binding-aware 3d molecule diffusion models. In *The Twelfth International Conference on Learning Representations* (2024).

34. Karnaukhov, V. K. *et al.* Structure-based prediction of t cell receptor recognition of unseen epitopes using tcren. *Nature Computational Science* **4**, 510–521 (2024).

35. Schneuing, A. *et al.* Structure-based drug design with equivariant diffusion models. *Nature Computational Science* **4**, 899–909 (2024).

36. Zhang, Z., Min, Y., Zheng, S. & Liu, Q. Molecule generation for target protein binding with structural motifs. In *The Eleventh International Conference on Learning Representations* (2023).

37. Lin, H. *et al.* Functional-group-based diffusion for pocket-specific molecule generation and elaboration. *Advances Neural Information Processing Systems* **36**, 34603–34626 (2023).

38. Jiang, Y. *et al.* Pocketflow is a data-and-knowledge-driven structure-based molecular generative model. *Nature Machine Intelligence* **6**, 326–337 (2024).

39. Zhang, Z. & Liu, Q. Learning subpocket prototypes for generalizable structure-based drug design. In *International Conference on Machine Learning*, 41382–41398 (PMLR, 2023).

40. Ragoza, M., Masuda, T. & Koes, D. R. Generating 3d molecules conditional on receptor binding sites with deep generative models. *Chemical science* **13**, 2701–2713 (2022).

41. Pinheiro, P. O., Jamasb, A., Mahmood, O., Sresht, V. & Saremi, S. Structure-based drug design by denoising voxel grids. *arXiv preprint arXiv:2405.03961* (2024).

42. Kong, X., Jia, Y., Huang, W. & Liu, Y. Full-atom peptide design with geometric latent diffusion. *Advances Neural Information Processing Systems* **37**, 74808–74839 (2024).

43. Kong, X. *et al.* Unimomo: Unified generative modeling of 3d molecules for de novo binder design. *arXiv preprint arXiv:2503.19300* (2025).

44. Peng, X. *et al.* Atom-level generative foundation model for molecular interaction with pockets. *bioRxiv* 2024–10 (2024).

45. Gao, Z. *et al.* Uniif: Unified molecule inverse folding. *Advances Neural Information Processing Systems* **37**, 135843–135860 (2024).

46. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature* **630**, 493–500 (2024).

47. Schmidt, F., Brennan, R. J., Jenkinson, S. & Valentin, J.-P. Shaping secondary pharmacology panels of the future: evolving target selection criteria for safety panels. *Nature Reviews Drug Discovery* 1–3 (2025).

48. Lounkine, E. *et al.* Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **486**, 361–367 (2012).

49. Reinecke, M. *et al.* Chemical proteomics reveals the target landscape of 1,000 kinase inhibitors. *Nature Chemical Biology* **20**, 577–585 (2024).

50. Davis, M. I. *et al.* Comprehensive analysis of kinase inhibitor selectivity. *Nature biotechnology* **29**, 1046–1051 (2011).

51. Liu, J. *et al.* In silico off-target profiling for enhanced drug safety assessment. *Acta Pharmaceutica Sinica B* **14**, 2927–2941 (2024).

52. Wei, L., Ye, X., Sakurai, T., Mu, Z. & Wei, L. Toxibtl: prediction of peptide toxicity based on information bottleneck and transfer learning. *Bioinformatics* **38**, 1514–1524 (2022).

53. Atz, K. *et al.* Prospective de novo drug design with deep interactome learning. *Nature Communications* **15**, 3408 (2024).

54. Zheng, L. *et al.* Improving protein–ligand docking and screening accuracies by incorporating a scoring function correction term. *Briefings Bioinformatics* **23**, bbac051 (2022).

55. Jing, B., Jaakkola, T. & Berger, B. Equivariant scalar fields for molecular docking with fast fourier transforms. *arXiv preprint arXiv:2312.04323* (2023).

56. Gao, B. *et al.* Rethinking specificity in sbdd: Leveraging delta score and energy-guided diffusion. *arXiv preprint arXiv:2403.12987* (2024).

57. Huggins, D. J., Sherman, W. & Tidor, B. Rational approaches to improving selectivity in drug design. *Journal medicinal chemistry* **55**, 1424–1444 (2012).

58. Kong, X., Huang, W., Tan, Z. & Liu, Y. Molecule generation by principal subgraph mining and assembling. *Advances Neural Information Processing Systems* **35**, 2550–2563 (2022).

59. Vincent, P. *et al.* Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal machine learning research* **11** (2010).

60. Jiao, R., Kong, X., Yu, Z., Huang, W. & Liu, Y. Equivariant pretrained transformer for unified geometric learning on multi-domain 3d molecules. In *ICLR 2024 Workshop on Generative and Experimental Perspectives for Biomolecular Design* (2024).

61. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

62. Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M. & Le, M. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747* (2022).

63. Sterling, T. & Irwin, J. J. Zinc 15–ligand discovery for everyone. *Journal chemical information modeling* **55**, 2324–2337 (2015).

64. Gaulton, A. *et al.* Chembl: a large-scale bioactivity database for drug discovery. *Nucleic acids research* **40**, D1100–D1107 (2012).

65. Burley, S. K. *et al.* Rcsb protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic acids research* **47**, D464–D474 (2019).

66. Wen, Z., He, J., Tao, H. & Huang, S.-Y. Pepbdb: a comprehensive structural database of biological peptide–protein interactions. *Bioinformatics* **35**, 175–177 (2019).

67. Wang, R., Fang, X., Lu, Y., Yang, C.-Y. & Wang, S. The pdbbind database: methodologies and updates. *Journal medicinal chemistry* **48**, 4111–4119 (2005).

68. Hu, L., Benson, M. L., Smith, R. D., Lerner, M. G. & Carlson, H. A. Binding moad (mother of all databases). *Proteins: Structure, Function, Bioinformatics* **60**, 333–340 (2005).

69. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proceedings National Academy Sciences* **89**, 10915–10919, DOI: 10.1073/pnas.89.22.10915 (1992).

70. Luo, S. *et al.* Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. *Advances Neural Information Processing Systems* **35**, 9754–9767 (2022).

71. Berrada, T. *et al.* Boosting latent diffusion with perceptual objectives. In *The Thirteenth International Conference on Learning Representations* (2025).

72. Tsaban, T. *et al.* Harnessing protein folding neural networks for peptide–protein docking. *Nature communications* **13**, 176 (2022).

73. Francoeur, P. G. *et al.* Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal chemical information modeling* **60**, 4200–4215 (2020).

74. Joeres, R., Blumenthal, D. B. & Kalinina, O. V. Data splitting to avoid information leakage with datasail. *Nature Communications* **16**, 3337, DOI: 10.1038/s41467-025-58606-8 (2025).

75. Yu, Y. *et al.* Uni-dock: Gpu-accelerated docking enables ultralarge virtual screening. *Journal chemical theory computation* **19**, 3336–3345 (2023).

76. Trott, O. & Olson, A. J. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal computational chemistry* **31**, 455–461 (2010).

77. Lin, H. *et al.* Cbgbench: fill in the blank of protein-molecule complex binding graph. *arXiv preprint arXiv:2406.10840* (2024).

78. Schake, P., Bolz, S. N., Linnemann, K. & Schroeder, M. Plip 2025: introducing protein–protein interactions to the protein–ligand interaction profiler. *Nucleic Acids Research* gkaf361 (2025).

79. Qu, Y. *et al.* Molcraft: Structure-based drug design in continuous parameter space. *arXiv preprint arXiv:2404.12141* (2024).