APOFLOWMOL: GEOMETRY-CONSTRAINED FLOW MATCHING FOR 3D MOLECULAR GENERATION IN FLEXIBLE PROTEIN POCKETS

Anonymous authors
Paper under double-blind review

ABSTRACT

Deep generative models have advanced structure-based drug design (SBDD), but most methods treat protein binding pockets as rigid, overlooking ligand-associated conformational changes. Existing flexible-pocket generators address this limitation, yet residue-wise coordinate or frame updates do not inherently preserve peptide-chain geometry. We introduce ApoFlowMol, a geometry-constrained full-atom flow-matching framework that jointly generates 3D ligands and adapts apo pockets to the evolving ligand state. ApoFlowMol employs a hierarchical SE(3)-equivariant complex network to model ligand coordinates, atom and bond identities, and protein conformational changes. Protein motion is represented either by fixed-backbone local torsion updates or by chain-linked backbone-and-side-chain kinematics. Both representations preserve sequence-defined peptide links recognized among modeled pocket residues: the former fixes the apo N–C_α–C trace, whereas the latter permits backbone motion while preserving backbone bond lengths and peptide connectivity within the pocket by construction throughout sampling. ApoFlowMol generates high-affinity molecules with performance comparable to leading flexible-pocket generative models. The resulting adapted pocket conformations provide more effective conditioning than unadapted apo pockets for established fixed-pocket SBDD methods. Surprisingly, pocket conformations selected through lead-guided side-chain-flexible docking similarly enable a fixed-pocket generator to achieve performance comparable to end-to-end flexible-pocket models. Together, these results highlight the potential of geometry-constrained pocket adaptation for flexible-pocket molecular generation and for improving existing fixed-pocket SBDD methods.

1 INTRODUCTION

Structure-based drug design (SBDD) uses three-dimensional target structures to guide the discovery of compatible ligands (Anderson, 2003). Recent E(3)- or SE(3)-equivariant diffusion and flow models jointly generate ligand coordinates and categorical atom types conditioned on protein pockets, while some additionally treat bond orders as explicit generative variables (Schneuing et al., 2024; Guan et al., 2023a;b; Cremer et al., 2026). However, these methods keep the receptor fixed throughout generation and therefore cannot capture the coupled evolution of the ligand and apo pocket. Recent flexible-pocket generative models address this limitation by allowing ligand and protein states to evolve together, with protein motion parameterized by residue-level frames, atom-level Cartesian coordinates, or side-chain torsions (Zhang et al., 2024; Zhou et al., 2025; Zheng et al., 2026; Wang et al., 2026). This progress shifts the challenge from whether the pocket is allowed to move to how its motion is parameterized. In representative methods that permit backbone motion, backbone conformations are represented using residue-wise rigid-body frames or Cartesian states. Such parameterizations do not explicitly enforce sequence-defined inter-residue peptide connectivity, including the C_i–N_{i+1} linkage, as a structural invariant throughout sampling (Zhang et al., 2024; Zhou et al., 2025; Zheng et al., 2026). SE(3) equivariance guarantees consistency under global rigid transformations, but does not by itself preserve the local covalent geometry connecting neighboring residues.

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.

To address this gap, we introduce **ApoFlowMol**, a geometry-constrained ligand–pocket co-flow framework that generates ligand coordinates, atom types, and bond types while adapting an apo pocket to the evolving ligand state. At every integration step, a hierarchical SE(3)-equivariant network re-encodes the kinematically reconstructed protein–ligand complex and predicts clean endpoints for continuous ligand and protein states, together with atom- and bond-class probabilities. SDE variants additionally predict score terms for the continuous states, while categorical variables follow a non-absorbing discrete flow. ApoFlowMol supports two complementary protein representations. The fixed-backbone representation preserves the apo N–C_α–C trace while updating carbonyl χ_0 and chemically defined side-chain $\chi_{1:4}$ torsions, whereas the chain-linked representation uses internal-coordinate kinematics to update both the backbone and side chains while preserving template backbone bond lengths and recognized sequence-adjacent peptide C_i–N_{i+1} bonds throughout sampling. Combining the two protein representations with deterministic ODE and stochastic SDE formulations yields four primary variants. We highlight our main contributions as follows:

- We introduce ApoFlowMol, a hierarchical $\mathrm{SE}(3)$-equivariant co-flow that jointly generates ligand geometry and chemistry while adapting an apo pocket.
- We develop complementary χ-only and chain-linked protein representations; the latter permits backbone motion while preserving backbone bond lengths and recognized peptide connectivity by construction.
- We evaluate pocket adaptation through direct generation, pocket reuse, and topology-aware peptide-continuity analysis, and show that both learned adaptation and lead-guided flexible docking can improve fixed-pocket ligand generation.

2 Related Work

2.1 Structure-Based Drug Design

Structure-based generative models construct 3D ligands conditioned on target protein pockets. Pocket2Mol autoregressively predicts atom types, coordinates, and bond types conditioned on a fixed protein pocket (Peng et al., 2022). DiffSBDD and TargetDiff instead employ non-autoregressive diffusion to jointly generate ligand coordinates and atom types while treating the target protein pocket as fixed conditioning context during generation (Schneuing et al., 2024; Guan et al., 2023a). DecompDiff decomposes each ligand into a scaffold and arms, defines decomposed priors over these ligand components, and explicitly diffuses ligand bonds; the receptor itself remains fixed throughout generation (Guan et al., 2023b). FLOWR provides a unified flow-matching framework for fixed-pocket ligand design, supporting structure-aware *de novo* generation as well as interaction-conditioned and fragment-based generation (Cremer et al., 2026). Collectively, these methods advance ligand generation but do not model ligand generation and apo-to-holo pocket adaptation as a coupled process.

2.2 Flexible Protein Modeling

Deep learning-based flexible docking methods such as DynamicBind (Lu et al., 2024) and FlowDock (Morehead & Cheng, 2025) predict ligand binding poses together with protein conformational adaptation for given ligand molecules; in contrast, *de novo* SBDD must additionally generate the ligand chemistry and three-dimensional structure. Flexible-pocket generative models address this setting by allowing ligand and protein states to evolve jointly. FlexSBDD represents each residue using a C_α coordinate, an $\mathrm{SO}(3)$ frame orientation, and side-chain χ angles (Zhang et al., 2024). DynamicFlow jointly models ligand coordinates, atom types, and bond types with protein residue-frame translations and rotations and torsional degrees of freedom, using deterministic and stochastic flow formulations (Zhou et al., 2025). Apo2Mol constructs 24,601 experimentally resolved apo–holo–ligand triplets from PLINDER (Durairaj et al., 2024) and jointly generates ligand coordinates and atom types while refining the pocket through per-residue translations, quaternion rotations, and side-chain χ updates (Zheng et al., 2026). DrugFlow samples ligand variables together with side-chain torsions while keeping the protein backbone fixed, whereas YuelDesign adopts a full-atom representation to jointly refine ligand and pocket coordinates while anchoring protein C_α atoms (Schneuing et al., 2025; Wang et al., 2026). While existing flexible-pocket generators incorporate

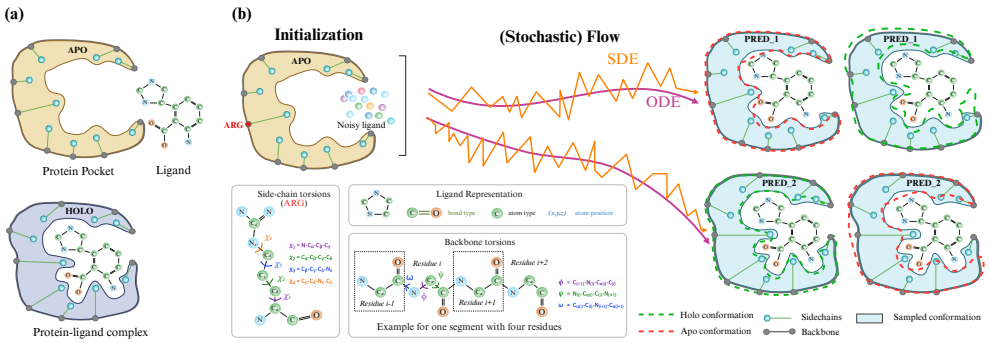


Figure 1: Overview of ApoFlowMol. (a) Flexible-pocket generation task. (b) ODE/SDE dynamics jointly update ligand coordinates, atom/bond types, and pocket geometry. ApoFlowMol-χ (PRED_1) updates $\chi_{0:4}$ torsions with a fixed apo N–C_α–C trace; ApoFlowMol-Link (PRED_2) additionally enables chain-linked backbone motion while preserving recognized peptide links. Dashed contours: apo (red) and holo (green).

protein motion, representative backbone-moving methods do not explicitly encode inter-residue peptide connectivity as a sampling invariant (Zhang et al., 2024; Zhou et al., 2025; Zheng et al., 2026; Wang et al., 2026). Unlike these approaches, ApoFlowMol compares a chemical-torsion state, which retains the apo N–C_α–C trace exactly, with a chain-linked state that permits backbone motion while preserving recognized peptide connectivity throughout sampling.

3 Method

In this section, we present our method, named ApoFlowMol, as illustrated in Fig. 1. We first formulate flexible-pocket molecular generation and introduce two pocket-motion parameterizations: a fixed N–C_α–C backbone trace with carbonyl χ_0 and side-chain $\chi_{1:4}$ updates (ApoFlowMol-χ), and chain-linked backbone-and-side-chain motion (ApoFlowMol-Link). We then present the shared ligand–pocket co-flow and hierarchical equivariant predictor, followed by the ODE/SDE training objectives and sampling procedures.

3.1 Data Preparation and Design Rationale

Task. We use the Apo2Mol benchmark, comprising 24,601 experimentally resolved apo–holo–ligand triplets curated from PLINDER, and adopt its original chronological split of 23,052/1,071/478 for training, validation, and testing, respectively (Durairaj et al., 2024; Zheng et al., 2026). Our structural parser retains 22,985, 1,071, and 478 systems in the training, validation, and test sets, respectively (Appendix A.1). Each system provides an apo pocket $\mathcal{P}^A$, a corresponding holo pocket $\mathcal{P}^H$, and a bound ligand $\mathcal{L}^H = (\mathbf{X}^H, \mathbf{A}^H, \mathbf{E}^H)$ containing coordinates, atom identities, and undirected bond identities. For pocket-motion parameterization m, ApoFlowMol models

$$p_\theta(N, \mathcal{L}_1, \mathcal{S}_1^m \mid \mathcal{P}^A) = p_\theta(N \mid \mathcal{P}^A)\, p_\theta(\mathcal{L}_1, \mathcal{S}_1^m \mid \mathcal{P}^A, N),\ \mathcal{P}_1 = K_m(\mathcal{P}^A, \mathcal{S}_1^m),\ m \in \{\chi, \text{Link}\} \tag{1}$$

where N is the ligand size and K_m reconstructs the full-atom pocket from the apo template and terminal motion state. Given a predefined binding site, generation requires only the corresponding apo pocket $\mathcal{P}^A$; holo coordinates and ligand identities are used only for supervision and evaluation. For controlled comparisons, the reference ligand atom count is supplied to remove ligand-size variation, but no other reference ligand information is used. For computationally intensive repeated generation, we cluster ESM-2 embeddings of the full 478 test set and select the observed sample nearest each of 100 cluster centroids (Lin et al., 2023). Details are presented in Appendix A.2.

Paired apo–holo analysis shows predominantly modest residue-level displacements with a long-motion tail, together with frequent apo-specific steric occlusion after calibration against matched holo contacts (Appendices A.3 and A.4). These observations motivate ApoFlowMol-χ, which fixes the apo N–C_α–C trace while updating carbonyl χ_0 and side-chain $\chi_{1:4}$ torsions. To test the same

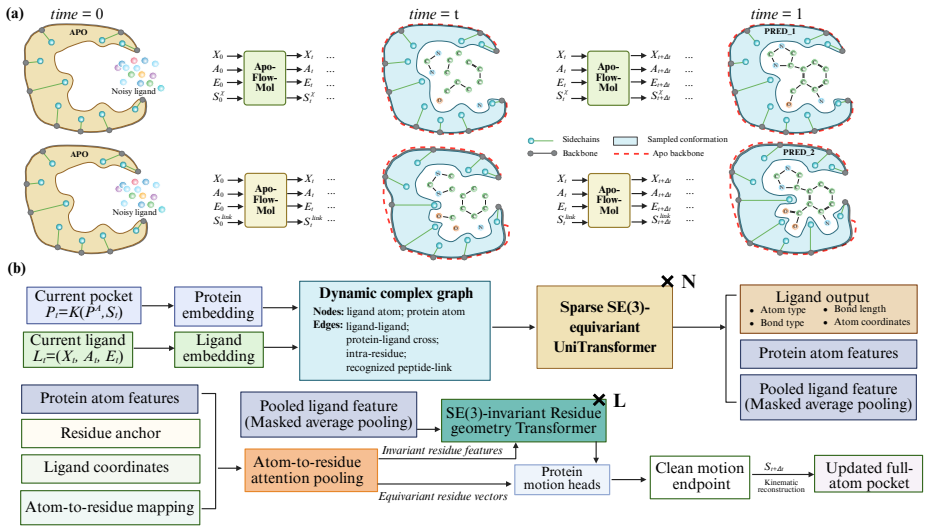


Figure 2: ApoFlowMol generation and architecture. (a) Ligand–pocket co-generation with χ-only (PRED_1) or chain-linked (PRED_2) protein motion under ODE/SDE dynamics. (b) Hierarchical full-atom and residue-level architecture for ligand and protein-motion prediction.

premise independently, an apo-site/ligand ranker selects leads and SMINA optimizes their poses and torsions together with at most ten nearby side chains (Trott & Olson, 2010; Koes et al., 2013). The adapted pockets then condition FLOWR (Cremer et al., 2026) to evaluate whether local relaxation improves fixed-pocket generation. This experiment is evaluation-only and independent of ApoFlowMol (Appendix A.5).

We also audit Apo2Mol and YuelDesign pockets generated using their released implementations and pretrained weights. Their sampling parameterizations do not preserve sequence-defined C_i–N_{i+1} links as kinematic invariants (Zheng et al., 2026; Wang et al., 2026). In our topology-defined audit, most evaluated pockets contain at least one expected link outside the 1.33 ± 0.20 Å criterion (Appendix A.6). This motivates ApoFlowMol-Link, which permits chain-linked backbone and χ-torsion adaptation while preserving recognized peptide connectivity by construction. Crossing the χ and Link parameterizations with deterministic ODE and stochastic SDE dynamics yields four separately trained models.

Kinematically reachable supervision. We enforce one-to-one, residue-identity-preserving apo–holo correspondence. Using matched protein C_α atoms, we estimate the global holo-to-apo rigid transform with the Kabsch least-squares solution (Kabsch, 1976) inside a deterministic iterative inlier-refit procedure. The same transform is then applied to all holo protein and ligand coordinates. Unmatched or spatially inconsistent residues remain available as apo context but are excluded from correspondence-based protein-motion and coordinate supervision. The masks used by individual interaction losses are described in Appendix B. For pocket-motion parameterization m, we fit a kinematically reachable terminal state:

$$\mathcal{S}_1^{m,*} = \mathrm{Fit}_m(\mathcal{P}^A, \mathcal{P}^H), \qquad \widetilde{\mathcal{P}}_m^H = K_m(\mathcal{P}^A, \mathcal{S}_1^{m,*}). \tag{2}$$

Here $\widetilde{\mathcal{P}}_m^H$ is the reachable reconstruction fitted to the holo pocket. Motion endpoints follow this fit; Link coordinate supervision uses $\widetilde{\mathcal{P}}_{\mathrm{Link}}^H$, whereas the χ coordinate auxiliary uses aligned holo atoms. Alignment, symmetry, masks, and fitting are detailed in Appendix B.

3.2 ApoFlowMol Framework

$$\mathcal{L}_t = (\mathbf{X}_t, \mathbf{A}_t, \mathbf{E}_t), \qquad \mathcal{P}_t = K_m(\mathcal{P}^A, \mathcal{S}_t^m), \qquad \mathcal{Z}_t = (\mathcal{L}_t, \mathcal{S}_t^m), \tag{3}$$

where $m \in \{\chi, \mathrm{Link}\}$ denotes the configured pocket-motion parameterization. The corresponding protein states are

$$\mathcal{S}_t^\chi = \{\Delta\chi_{0,r}(t), \Delta\chi_{r,k}(t)\}\,, \tag{4}$$

$$\mathcal{S}_t^{\text{Link}} = \left\{ \big(\mathbf{t}_s(t), \mathbf{R}_s(t)\big), \Delta\phi_r(t), \Delta\psi_r(t), \Delta\omega_r(t), \Delta\alpha_{r,j}(t), \Delta\chi_{0,r}(t), \Delta\chi_{r,k}(t) \right\}, \tag{5}$$

where s, r, j, and k index continuous peptide segments, residues, backbone bond angles, and residue-dependent side-chain torsions, respectively. Each segment-root transform is defined relative to its apo structure: an atom $\mathbf{x}$ in segment s is transformed as $\mathbf{a}_s^A + \mathbf{t}_s(t) + \mathbf{R}_s(t)(\mathbf{x} - \mathbf{a}_s^A)$, where $\mathbf{a}_s^A$ is the fixed apo anchor, with $\mathbf{t}_s(0) = \mathbf{0}$ and $\mathbf{R}_s(0) = \mathbf{I}$. Here, χ_0 controls rotation of the backbone carbonyl group about the C_α–C bond, while $\chi_{1:4}$ denote the standard side-chain torsions.

Fig. 2(a) depicts the recurrent evolution of the ligand together with either protein state. ApoFlowMol-χ keeps all apo N, C_α, and C coordinates fixed and updates carbonyl χ_0 and chemically defined side-chain $\chi_{1:4}$ torsions (Schneuing et al., 2025). ApoFlowMol-Link additionally uses one root $\mathrm{SE}(3)$ transform per continuous peptide segment and linked $\phi/\psi/\omega$, with predicted backbone-angle changes constrained to amino-acid-specific ranges. The key distinction is that backbone motion is propagated through a connected peptide chain rather than through independently moving residue frames, while carbonyl and side-chain flexibility remains governed by $\chi_{0:4}$. For the Link model, a NeRF-style chain reconstruction (Parsons et al., 2005; AlQuraishi, 2019) maps the motion state to full-atom coordinates while reusing apo-template covalent lengths. Let $\mathcal{E}_{\text{bb}}^A$ contain the recognized intra-residue N–C_α and C_α–C bonds and inter-residue C_i–N_{i+1} peptide bonds. Then

$$\|\mathbf{x}_i(t) - \mathbf{x}_j(t)\|_2 = d_{ij}^A, \qquad (i,j) \in \mathcal{E}_{\text{bb}}^A, \quad \forall\, t \in [0,1]. \tag{6}$$

Thus, recognized backbone edges retain their apo-template lengths throughout sampling without post-hoc projection. This guarantee does not extend to unrecognized covalent edges, peptide-plane geometry, or nonbonded contacts.

Continuous and categorical dynamics. At $t = 0$, ligand coordinates are sampled from a pocket-centered Gaussian, atom and bond identities from uniform priors, and protein motion from the identity state. During training, we shorten exchangeable ligand-coordinate paths using a FLOWR-style equivariant-OT coupling (Klein et al., 2023; Cremer et al., 2026). A Hungarian assignment is followed by a rotation-only Kabsch alignment about the apo-pocket center $\mathbf{c}_P$:

$$\pi^* = \arg\min_{\pi \in \Pi_N} \sum_i \left\| (\mathbf{x}_i^H - \mathbf{c}_P) - (\mathbf{x}_{0,\pi(i)} - \mathbf{c}_P) \right\|_2^2, \tag{7}$$

$$\mathbf{R}^* = \arg\min_{\mathbf{R} \in \mathrm{SO}(3)} \sum_i \left\| (\mathbf{x}_i^H - \mathbf{c}_P) - \mathbf{R}(\mathbf{x}_{0,\pi^*(i)} - \mathbf{c}_P) \right\|_2^2. \tag{8}$$

This target-dependent coupling is used only to construct training paths. For each Euclidean or wrapped-periodic continuous modality u, we define

$$z_t^u = z_0^u + t\,\Delta_u(z_0^u, z_1^u) + \mathbf{1}_{\text{SDE}}\,\sigma_u \sqrt{t(1-t)}\,\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{9}$$

where periodic variables use wrapped displacements. Rotations use the analogous group-valued geodesic bridge on $\mathrm{SO}(3)$, as detailed in Appendix D. A predicted clean endpoint defines the endpoint-induced field

$$q_\theta^u(\mathcal{Z}_t, t) = \frac{\mathrm{Log}_{z_t^u}(\widehat{z}_{1,\theta}^u)}{1-t}. \tag{10}$$

For the unconstrained Euclidean bridge in Eq. 9, the population endpoint field under squared endpoint regression differs from the probability velocity by $\frac{1}{2}\sigma_u^2 s^u$. This motivates the following sampling equations; the periodic and rotational implementations are described in Appendix D.

$$\begin{cases} dz_t^u = q_\theta^u\,dt, & \text{ODE}, \\ dz_t^u = \left[q_\theta^u + \frac{1}{2}\left(g_u^2 - \sigma_u^2\right) s_\theta^u \right] dt + g_u\,dW_t^u, & \text{SDE}, \end{cases} \qquad g_u^2 = 2\sigma_u^2(1-t). \tag{11}$$

Atom and bond identities follow the FLOWR-style categorical path (Cremer et al., 2026)

$$q_t(c_t \mid c_1) = t\,\delta_{c_1}(c_t) + \frac{1-t}{K_c}. \tag{12}$$

A non-absorbing corrector allows categorical predictions to be revised across steps, while bond updates remain symmetric. For soft self-conditioning, training uses endpoint probabilities from a stop-gradient preliminary prediction at the same noisy state, whereas sampling reuses those retained from the preceding integration step. Further implementation details and modality-specific noise scales are provided in Appendix D.

Hierarchical predictor. Figure 2 shows the atom- and residue-level predictor evaluated at each step. Inspired by pharmacophore conditioning in DiffPharma (Sako et al., 2026), we derive typed interaction sites solely from apo residue and atom identities; their coordinates follow the reconstructed pocket $\mathcal{P}_t$. Bidirectional ligand–site attention augments the atom embeddings before a sparse $\mathrm{SE}(3)$-equivariant UniTransformer exchanges ligand–protein messages (Guan et al., 2023a; Zheng et al., 2026). Only ligand message coordinates are updated equivariantly; protein coordinates remain fixed to $K_m(\mathcal{P}^A, \mathcal{S}_t^m)$ within the encoder. Periodic internal states enter as invariant embeddings, whereas segment-root states enter through reconstructed coordinates. Atom-to-residue pooling yields scalar residue features and equivariant vectors from anchor-relative geometry and a soft ligand-contact displacement. After adding ligand context and categorical confidence, the residue Transformer processes only invariant features and distance/link biases; equivariant vectors bypass it to parameterize the segment-root heads. Overall,

$$\left(\widehat{\mathbf{X}}_1, p_\theta(\mathbf{A}_1), p_\theta(\mathbf{E}_1), \widehat{\mathbf{D}}_1, \widehat{\mathcal{S}}_1^m, \mathbf{s}_\theta\right) = F_\theta(\mathcal{Z}_t, \mathcal{P}^A, t; \mathcal{C}_t^{\mathrm{sc}}), \qquad p_\theta(N \mid \mathcal{P}^A) = G_\theta(\mathcal{P}^A), \tag{13}$$

where $\mathcal{C}_t^{\mathrm{sc}}$ denotes the soft categorical condition available to the current prediction, $\widehat{\mathbf{D}}_1$ is an auxiliary symmetric predicted-bond length endpoint, and $\mathbf{s}_\theta$ is present only in SDE models. Each checkpoint uses one protein head, producing four variants: ApoFlowMol-χ-ODE/SDE and ApoFlowMol-Link-ODE/SDE.

3.3 Training Objective

For each training complex, a shared $t \sim \mathcal{U}(0,1)$ defines the current ligand and protein states. Both ODE and SDE models are trained through clean-endpoint prediction. These endpoints define the endpoint-induced fields in Eq. 11, while SDE models additionally learn the conditional scores of the stochastic bridge. The same clean endpoints are used for reachable-coordinate and geometric auxiliary losses. We use cross-entropy losses for ligand size, atom identities, and bond identities; masked MSE or Huber losses for Euclidean variables; wrapped errors for torsions; and squared geodesic errors for rotations. We group the normalized terms as

$$\mathcal{L}_{\mathrm{lig}} = \lambda_X \mathcal{L}_X + \lambda_A \mathcal{L}_A + \lambda_E \mathcal{L}_E + \lambda_N \mathcal{L}_N, \quad \mathcal{L}_{\mathrm{phys}} = \mathcal{L}_{\mathrm{prot}}^m + \mathcal{L}_{\mathrm{geo}} + \lambda_{\mathrm{mot}} \mathcal{L}_{\mathrm{motion}} \tag{14}$$

where $\mathcal{L}_{\mathrm{prot}}^m$ supervises the active motion variables and the reachable pocket defined in Eq. 2, while $\mathcal{L}_{\mathrm{geo}}$ covers coordinate-based bonded distances, predicted bond-length endpoints, and bonded angles. The complex-level term $\mathcal{L}_{\mathrm{complex}}$ combines local full-complex geometry, dedicated protein–ligand cross distances, pharmacophore compatibility, and severe-clash penalties. Following the delayed interaction supervision of DynamicFlow (Zhou et al., 2025), we activate these complex-level terms only for $t > 0.65$, while ligand and pocket dynamics remain supervised throughout:

$$\mathcal{L} = \mathcal{L}_{\mathrm{lig}} + \mathcal{L}_{\mathrm{phys}} + \mathbf{1}[t > 0.65]\mathcal{L}_{\mathrm{complex}} + \lambda_s \mathbf{1}_{\mathrm{SDE}} \mathcal{L}_{\mathrm{score}}. \tag{15}$$

At inference, ligand and pocket states are synchronously advanced, kinematically reconstructed, and re-encoded for 56 steps. Exact component losses, weights, noise scales, optimization settings, and optional refinement are provided in Appendices E–F.

4 Experiments

4.1 Experimental setup

Dataset. We use the chronological Apo2Mol split of 23,052 training, 1,071 validation, and 478 test complexes (Zheng et al., 2026). Repeated generation is evaluated on a sequence-diverse 100-system subset selected from the held-out test set before inspecting model outputs (Appendices A.1–A.2).

Baseline methods. For direct ligand–pocket generation, we compare ApoFlowMol with Apo2Mol and YuelDesign. We evaluate the χ and Link representations with both ODE and SDE dynamics. The main tables report the primary configurations, while additional Link angle-control variants are provided in Appendix G.2.

Table 1: **Direct ligand–pocket generation on the sequence-diverse 100-system test subset.** Reference-apo and Reference-holo denote the native ligand evaluated in the matched apo and holo pockets, respectively; ApoFlowMol-Link denotes the Link-GR parameterization. Avg./Med. denote mean/median, and Vina is in $\mathrm{kcal\,mol^{-1}}$. The best and second-best generated-method results are bolded and underlined, respectively; ties are bolded. Reference controls and LogP are not ranked.

	Vina ↓		QED ↑		SA ↑		LogP		Lipinski ↑		HA ↑
Method	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	(%)
Reference-apo	-6.23	-6.10	0.60	0.61	0.77	0.79	2.05	2.48	4.69	5.00	–
Reference-holo	-7.04	-6.80	0.60	0.61	0.77	0.79	2.05	2.48	4.69	5.00	–
Apo2Mol	<u>-7.74</u>	<u>-7.90</u>	**0.54**	**0.54**	**0.54**	**0.56**	3.60	3.67	4.46	**5.00**	67
YuelDesign	-5.60	-5.60	0.21	0.18	0.50	0.50	-0.89	-0.80	3.20	3.00	19
ApoFlowMol-χ-ODE	-6.97	-7.15	0.43	0.42	<u>0.53</u>	<u>0.53</u>	0.52	0.43	4.55	**5.00**	57
ApoFlowMol-χ-SDE	-7.25	-7.15	0.46	0.42	0.48	0.47	0.49	0.33	<u>4.56</u>	**5.00**	65
ApoFlowMol-Link-ODE	-7.73	-7.65	<u>0.48</u>	<u>0.47</u>	0.48	0.47	0.28	0.57	**4.59**	**5.00**	<u>77</u>
ApoFlowMol-Link-SDE	**-7.90**	**-8.10**	0.39	0.37	0.32	0.28	-1.51	-1.52	4.35	**5.00**	**79**

Evaluation metrics. Following Apo2Mol (Zheng et al., 2026), we report the same six categories of ligand-quality metrics: 1) **binding affinity**, measured by the minimum successful Vina-GPU docking score; 2) **QED**, measuring drug-likeness; 3) **SA Score**, normalized synthetic accessibility, where a higher value indicates easier synthesis; 4) **LogP**, measuring octanol–water partitioning; 5) **Lipinski**, the number of five drug-likeness criteria satisfied by the molecule; and 6) **High Affinity** (HA), the percentage of selected ligands whose docking score is strictly lower than that of the matched reference ligand in its holo pocket. All metrics except HA are summarized by both their mean and median.

Evaluation protocol. All ApoFlowMol configurations use the reference ligand size and 56 synchronous ligand–pocket updates. For direct generation, each method contributes five ligand–pocket candidates per system, and the candidate with the lowest successful Vina-GPU score is selected. For pocket reuse, the selected pocket is fixed and used to condition FLOWR (Cremer et al., 2026); FLOWR-n selects the lowest successful docking score among the first n saved ligands. Docking uses Vina-GPU 2.1 with a 20 Å cubic box, search depth 1, and archived seeds (see Appendix G for details).

4.2 Joint ligand–pocket generation

ApoFlowMol-Link-ODE and ApoFlowMol-Link-SDE achieve direct Avg./Med. Vina scores of $-7.73/-7.65$ and $-7.90/-8.10\,\mathrm{kcal\,mol^{-1}}$, with corresponding HA rates of 77% and 79%, respectively (Table 1). Both configurations have higher HA point estimates than Apo2Mol's 67%. We additionally explored backbone-update schemes with different angle-activation and motion constraints. For concise presentation, the main text reports one Link parameterization selected from this exploratory comparison; definitions and complete results for all variants are provided in Appendix G.2. Among the eight ApoFlowMol configurations, ApoFlowMol-Link-SDE yields the most favorable mean Vina score, although its mean QED and SA are lower than those of ApoFlowMol-Link-ODE ($0.39/0.32$ versus $0.48/0.48$). The paired SDE-minus-ODE difference in mean Vina score is $-0.17\,\mathrm{kcal\,mol^{-1}}$ (95% bootstrap CI $[-0.43, 0.10]$); because the interval includes zero, the observed advantage of SDE is not statistically resolved by this paired analysis. The two ApoFlowMol-Link configurations preserve near-ideal peptide-link geometry, with a median per-output C–N MAE of approximately 0.006 Å, compared with 0.757 Å for Apo2Mol and 0.997 Å for YuelDesign (Figure 3a). Moreover, every evaluated ApoFlowMol-Link C–N distance deviates by less than 0.10 Å from the reference value of 1.33 Å.

Table 2: **Pocket reuse with FLOWR.** Avg./Med. denote mean/median Vina scores over 100 systems; HA is in percent. Bold and underline mark the best and second-best learned generated-pocket conditions, respectively; reference controls and the separately budgeted FlexDock analysis are excluded from this ranking. FlexDock reports the best single-library result; complete library-wise and exhaustive results are in Appendix G.4.

	FLOWR-5			FLOWR-100		
Conditioning pocket	Avg. ↓	Med. ↓	HA (%) ↑	Avg. ↓	Med. ↓	HA (%) ↑
Reference apo	-6.88	-6.50	48	-8.28	-8.10	81
Reference holo	-7.46	-7.10	62	-8.86	-8.80	95
Apo2Mol	-7.07	-6.95	47	-8.42	<u>-8.55</u>	82
YuelDesign	<u>-7.34</u>	<u>-7.05</u>	<u>56</u>	-8.51	-8.50	82
ApoFlowMol-χ-ODE	-6.99	-6.70	48	-8.28	-7.95	81
ApoFlowMol-χ-SDE	-6.87	-6.55	51	-8.24	-8.05	81
ApoFlowMol-Link-ODE	-7.18	-6.95	51	<u>-8.52</u>	-8.45	<u>89</u>
ApoFlowMol-Link-SDE	**-7.68**	**-7.60**	**68**	**-8.82**	**-8.80**	**90**
FlexDock (HighFid1920)	-7.73	-7.70	67	-8.55	-8.20	88

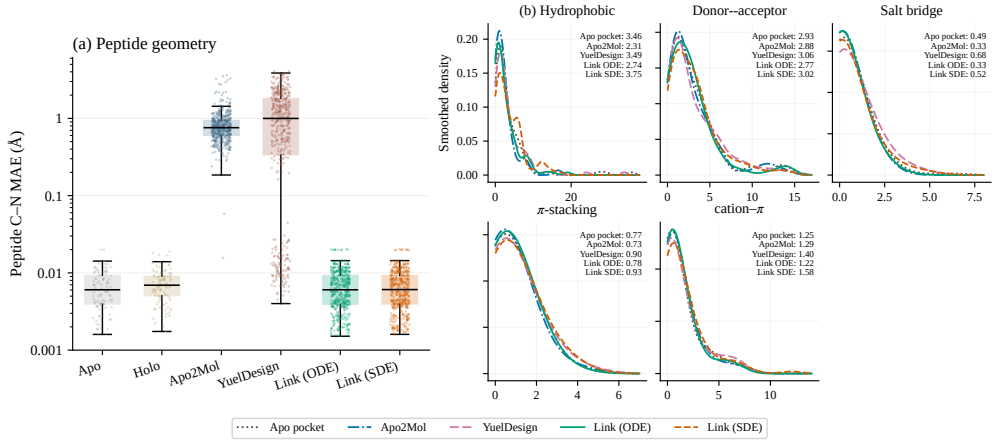
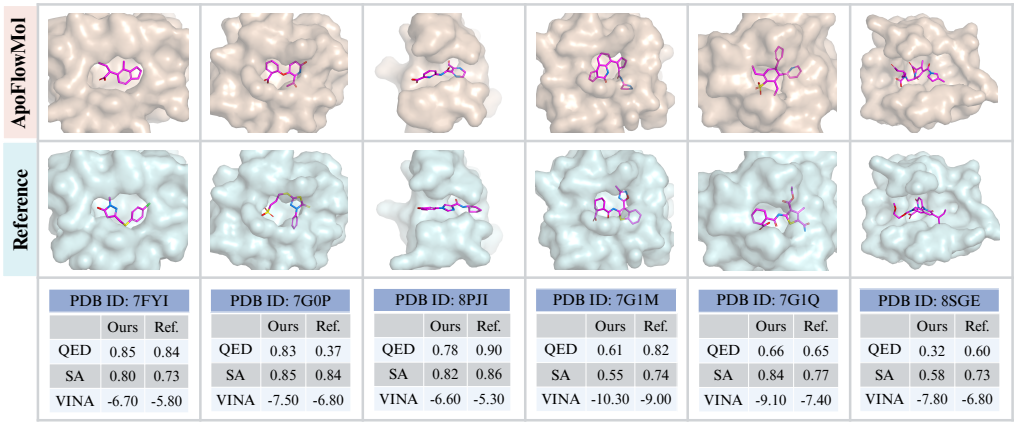


Figure 3: **Peptide geometry and interaction-count differences.** (a) Peptide C–N MAE for 100 apo/holo controls and 500 outputs per generated cohort on a logarithmic scale. Dots denote individual pockets; boxes show the median and interquartile range. (b) Absolute differences in five non-covalent interaction counts between FLOWR-100 winners under each pocket condition and the matched native-ligand holo complexes. Insets report mean absolute differences.

4.3 Pocket reuse and configuration trade-offs

Among generated-pocket conditions, ApoFlowMol-Link gives FLOWR-100 HA closest to holo conditioning. SDE achieves Avg./Med. Vina scores of $-8.82/-8.80$ with 90% HA, most closely approaching holo's $-8.86/-8.80$ and 95%; ODE gives $-8.52/-8.45$ and 89% (Table 2). Apo, Apo2Mol, and YuelDesign yield $-8.28/-8.10$, $-8.42/-8.55$, and $-8.51/-8.50$, with 81%, 82%, and 82% HA, respectively.

More unexpectedly, simple FlexDock preprocessing improves FLOWR by adjusting only selected side chains around a lead in the same apo structure, fixing the backbone, and freezing the adapted pocket for generation. All five lead libraries improve FLOWR-100 mean Vina scores and HA over apo (-8.55 to -8.40; 82–88% HA); HighFid1920 gives the best single-library mean (Avg./Med. $-8.55/-8.20$, 88% HA). Broader two-stage selection yields lower FLOWR-100 Avg./Med. Vina scores than holo conditioning ($-9.16/-9.00$ versus $-8.86/-8.80$), although HA remains lower (93% versus 95%). Exhaustive evaluation of all 2,399 candidate pockets across the test set reaches $-9.44/-9.30$ with 96% HA (Appendix G.4). These larger proposal-and-selection budgets do not establish matched-budget superiority. This suggests improving receptor inputs for established fixed-pocket SBDD through lead-guided docking while retaining the generator's architecture, parameters,



PDB ID: 7FYI		
	Ours	Ref.
QED	0.85	0.84
SA	0.80	0.73
VINA	-6.70	-5.80

PDB ID: 7G0P		
	Ours	Ref.
QED	0.83	0.37
SA	0.85	0.84
VINA	-7.50	-6.80

PDB ID: 8PJI		
	Ours	Ref.
QED	0.78	0.90
SA	0.82	0.86
VINA	-6.60	-5.30

PDB ID: 7G1M		
	Ours	Ref.
QED	0.61	0.82
SA	0.55	0.74
VINA	-10.30	-9.00

PDB ID: 7G1Q		
	Ours	Ref.
QED	0.66	0.65
SA	0.84	0.77
VINA	-9.10	-7.40

PDB ID: 8SGE		
	Ours	Ref.
QED	0.32	0.60
SA	0.58	0.73
VINA	-7.80	-6.80

Figure 4: Visualization of ligands and pockets generated by ApoFlowMol-Link, compared with the corresponding reference ligands and holo structures. QED, SA, and Vina docking scores are reported.

and fixed-pocket sampling without retraining. Explicit donor, acceptor, aromatic, or hydrophobic centers guide pocket selection and interpretation through specified interaction hypotheses.

ODE uses deterministic continuous-state updates, whereas SDE trains on noise-augmented paths and introduces stochastic perturbations during joint ligand–pocket sampling, potentially broadening conformational exploration. DynamicFlow reports higher holo-state coverage with SDE but lower minimum holo RMSD with ODE (Zhou et al., 2025). In our runs, SDE combines lower Vina point estimates with larger apo–generated motion (2.72 versus 0.23 Å) and higher motion JSD (0.899 versus 0.516), indicating a trade-off between docking scores and agreement with reference motion magnitudes. Alternative angle controls and all eight configurations are reported in Appendix G.2.

In Figure 3b (FLOWR-100), Apo2Mol and ApoFlowMol-Link-ODE reduce mean absolute hydrophobic-count deviations to 2.31 and 2.74, versus 3.46 for apo. ApoFlowMol-Link-ODE also has the lowest donor–acceptor contact-count deviation (2.77), with salt-bridge and cation–π deviations near apo; YuelDesign and ApoFlowMol-Link-SDE show broader deviations in several classes. Similar counts need not involve the same interacting residues. Figure 4 visualizes six illustrative ApoFlowMol-Link direct-generation examples alongside their matched reference ligand–holo-pocket pairs. Beyond these representative examples, the supplement provides paired comparisons of total typed-contact errors, per-type contact distributions including halogen contacts, and ligand C–C distance distributions.

5 Conclusion

ApoFlowMol couples ligand generation with geometry-constrained apo-pocket adaptation while preserving recognized peptide connectivity. ApoFlowMol-Link achieves favorable docking-based performance in both direct generation and pocket reuse, with ODE and SDE dynamics exhibiting different trade-offs in molecular properties and pocket motion. We further find that simple lead-guided side-chain-flexible docking provides effective receptor inputs for an unchanged FLOWR model. Under broader pocket proposal-and-selection budgets, these adapted pockets yield lower Vina scores than holo-conditioned FLOWR. Together, these findings support geometry-constrained joint generation and highlight a complementary strategy for improving established fixed-pocket SBDD workflows: interaction-guided receptor adaptation upstream of molecular generation, without modifying or retraining the generator.

AI use statement

We used large language models to assist with literature searches and identify relevant studies on flexible-pocket structure-based drug design and related receptor-adaptation methods. LLMs also

assisted with editing the manuscript for grammar and readability. We did not use generative AI tools for any other tasks requiring disclosure under the ICLR 2027 AI Policy. We reviewed all AI-assisted work and take responsibility for the final content of this work.

Reproducibility statement

We provide detailed descriptions of the data preparation, model architecture, training objectives, sampling procedures, hyperparameters, and evaluation protocols in the main text and appendices. Dataset construction, preprocessing, and test-subset selection are described in Appendices A.1 and A.2; correspondence and reachable-target construction in Appendix B; geometry-constrained protein states in Appendix C; co-flow and predictor details in Appendix D; objectives and optimization in Appendix E; and sampling and refinement in Appendix F. Evaluation protocols, model configurations, and additional results are provided in Appendices G and G.2. These details are intended to support reproduction of the reported experiments.

References

Mohammed AlQuraishi. Parallelized natural extension reference frame: Parallelized conversion from internal to cartesian coordinates. *Journal of Computational Chemistry*, 40(7):885–892, 2019. doi: 10.1002/jcc.25772.

Amy C. Anderson. The process of structure-based drug design. *Chemistry & Biology*, 10(9):787–797, 2003. doi: 10.1016/j.chembiol.2003.09.002.

A. Bondi. van der waals volumes and radii. *The Journal of Physical Chemistry*, 68(3):441–451, 1964. doi: 10.1021/j100785a001.

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*, 6:565–574, 2026. doi: 10.1038/s43588-026-00998-8.

Janani Durairaj, Yusuf Adeshina, Zhonglin Cao, Xuejin Zhang, Vladas Oleinikovas, Thomas Duignan, Zachary McClure, Xavier Robin, Gabriel Studer, Daniel Kovtun, Emanuele Rossi, Guoqing Zhou, Srimukh Veccham, Clemens Isert, Yuxing Peng, Prabindh Sundareson, Mehmet Akdel, Gabriele Corso, Hannes Stärk, Gerardo Tauriello, Zachary Carpenter, Michael Bronstein, Emine Kucukbenli, Torsten Schwede, and Luca Naef. PLINDER: The protein-ligand interactions dataset and evaluation resource. *bioRxiv*, 2024. doi: 10.1101/2024.07.17.603955. URL https://www.biorxiv.org/content/10.1101/2024.07.17.603955v3.

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 *The Eleventh International Conference on Learning Representations*, 2023a. URL https://openreview.net/forum?id=kJqXEPXMsE0.

Jiaqi Guan, Xiangxin Zhou, Yuwei Yang, Yu Bao, Jian Peng, Jianzhu Ma, Qiang Liu, Liang Wang, and Quanquan Gu. DecompDiff: Diffusion models with decomposed priors for structure-based drug design. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 11827–11846. PMLR, 23–29 Jul 2023b. URL https://proceedings.mlr.press/v202/guan23a.html.

Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976. doi: 10.1107/S0567739476001873.

Leon Klein, Andreas Krämer, and Frank Noé. Equivariant flow matching. In *Advances in Neural Information Processing Systems*, volume 36, pp. 59886–59910, 2023. doi: 10.52202/075280-2617. URL https://openreview.net/forum?id=eLH2NFOO1B.

David Ryan Koes, Matthew P. Baumgartner, and Carlos J. Camacho. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *Journal of Chemical Information and Modeling*, 53(8):1893–1904, 2013. doi: 10.1021/ci300604z.

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

Wei Lu, Jixian Zhang, Weifeng Huang, Ziqiao Zhang, Xiangyu Jia, Zhenyu Wang, Leilei Shi, Chengtao Li, Peter G. Wolynes, and Shuangjia Zheng. DynamicBind: Predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. *Nature Communications*, 15(1):1071, 2024. doi: 10.1038/s41467-024-45461-2. URL https://www.nature.com/articles/s41467-024-45461-2.

Alex Morehead and Jianlin Cheng. FlowDock: Geometric flow matching for generative protein-ligand docking and affinity prediction. *Bioinformatics*, 41(Supplement_1):i198–i206, 2025. doi: 10.1093/bioinformatics/btaf187. URL https://academic.oup.com/bioinformatics/article/41/Supplement_1/i198/8199366.

Jerod Parsons, J. Bradley Holmes, J. Mauricio Rojas, Jerry Tsai, and Charlie E. M. Strauss. Practical conversion from torsion space to cartesian space for in silico protein synthesis. *Journal of Computational Chemistry*, 26(10):1063–1068, 2005. doi: 10.1002/jcc.20237.

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2Mol: Efficient molecular sampling based on 3D protein pockets. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 17644–17655. PMLR, 2022. URL https://proceedings.mlr.press/v162/peng22b.html.

Masami Sako, Nobuaki Yasuo, and Masakazu Sekijima. Interaction-constrained 3D molecular generation using a diffusion model enables structure-based pharmacophore modeling for drug design. *npj Drug Discovery*, 3(1):8, 2026. doi: 10.1038/s44386-026-00040-x.

Arne Schneuing, Charles Harris, Yuanqi Du, Kieran Didi, Arian Jamasb, Ilia Igashov, Weitao Du, Carla Gomes, Tom L Blundell, Pietro Lio, Max Welling, Michael Bronstein, and Bruno Correia. Structure-based drug design with equivariant diffusion models. *Nature Computational Science*, 4 (12):899–909, 2024. doi: 10.1038/s43588-024-00737-x.

Arne Schneuing, Ilia Igashov, Adrian W. Dobbelstein, Thomas Castiglione, Michael Bronstein, and Bruno Correia. Multi-domain distribution learning for de novo drug design. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=g3VCIM94ke.

Oleg Trott and Arthur J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010. doi: 10.1002/jcc.21334.

Jian Wang, Dong Yan Zhang, Shreshty Budakoti, and Nikolay V. Dokholyan. A diffusion-based framework for designing molecules in flexible protein pockets. *Science Advances*, 12(15): eaeb7045, 2026. doi: 10.1126/sciadv.aeb7045.

Zaixi Zhang, Mengdi Wang, and Qi Liu. FlexSBDD: Structure-based drug design with flexible protein modeling. In *Advances in Neural Information Processing Systems*, volume 37, pp. 53918–53944. Curran Associates, Inc., 2024. doi: 10.52202/079017-1707. URL https://openreview.net/forum?id=4AB54h21qG.

Xinzhe Zheng, Shiyu Jiang, Gustavo Seabra, Chenglong Li, and Yanjun Li. Apo2Mol: 3D molecule generation via dynamic pocket-aware diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 1614–1622, 2026. doi: 10.1609/aaai.v40i2.37138. URL https://ojs.aaai.org/index.php/AAAI/article/view/37138.

Xiangxin Zhou, Yi Xiao, Haowei Lin, Xinheng He, Jiaqi Guan, Yang Wang, Qiang Liu, Feng Zhou, Liang Wang, and Jianzhu Ma. Integrating protein dynamics into structure-based drug design via full-atom stochastic flows. In *International Conference on Learning Representations*, volume 2025, pp. 18621–18649, 2025. URL https://openreview.net/forum?id=9qS3HzSDNv.

A Data and Design Evidence

A.1 Experimental apo–holo benchmark

We use the 24,601 experimental apo–holo–ligand triplets introduced by Apo2Mol and curated from PLINDER, together with the chronological 23,052/1,071/478 train/validation/test source split (Durairaj et al., 2024; Zheng et al., 2026). The holo pocket contains residues within 10 Å of the crystallographic ligand; the corresponding residues are extracted from the apo structure. Our parser accepts 22,985/1,071/478 systems. The 67 excluded training entries contain no residue with a complete N–C_α–C backbone. Source indices are retained. The released preprocessed LMDB used only for the whole-benchmark descriptive motion and steric analyses contains 24,500 records; it is distinct from both the source split and the ApoFlowMol training cache.

Generation receives the apo pocket and its sequence-defined topology. Holo coordinates and the reference ligand are used only to construct supervision and evaluation metrics. We supply the reference ligand atom count only in controlled benchmark comparisons; otherwise ligand size is predicted from the apo pocket.

A.2 Sequence-diverse 100-system evaluation subset

Repeated five-candidate, 56-step generation over eight configurations is expensive: χ, Link-G5, Link-GR, and Link-R, each with ODE and SDE dynamics. We therefore mean-pool ESM-2 representations of the 478 test pocket sequences, cluster them into 100 groups by K-means (seed 2021; ten initializations), and select the observed sample nearest each centroid (Lin et al., 2023). Selection is completed before inspecting any model output. This subset reduces sequence redundancy and covers more sequence clusters than an unstratified sample of the same size, but it is not a substitute for the complete test split and may miss within-cluster geometric or chemical variation. For the pocket-conditioning benchmark, five candidates per pocket are retained after refinement and candidate checks; Vina-GPU-based pocket selection is detailed in Appendix F.

A.3 Apo–holo motion analysis

For Table 3, we iterate over all 24,500 records in the released preprocessed LMDB and use their stored apo and holo coordinates, atom-row correspondence, and torsion values, without the additional alignment or symmetry canonicalization used for target construction in Appendix B.

For residue r, let $\mathcal{A}_{r,u}$ contain its backbone ($u = bb$) or side-chain ($u = sc$) heavy atoms, and let $\mathcal{K}_r$ contain its available chemical torsions. We compute

$$\begin{aligned} \mathrm{RMSD}_{r,u}^{A\to H} &= \left(|\mathcal{A}_{r,u}|^{-1} \sum_{a\in\mathcal{A}_{r,u}} \|\mathbf{x}_a^H - \mathbf{x}_a^A\|_2^2 \right)^{1/2}, \quad u \in \{bb, sc\}, \\ \Delta\chi_{\max,r} &= \max_{k\in\mathcal{K}_r} \left| \mathrm{wrap}_{2\pi}(\chi_{r,k}^H - \chi_{r,k}^A) \right|. \end{aligned} \tag{16}$$

This descriptive torsion statistic uses a uniform 2π period and retains symmetry-equivalent terminal-atom flips; the symmetry-aware training targets are defined separately in Appendix B. Table 3 shows predominantly modest residue-level displacements, together with a long-motion tail.

Table 3: Residue-level apo–holo motion across all 24,500 records in the released preprocessed LMDB. Backbone RMSD, side-chain RMSD, and maximum torsion change summarize 997,802, 900,914, and 820,639 eligible residues, respectively. Each eligible residue contributes one value; tail frequencies use the corresponding residue count as the denominator.

Quantity	Median	Mean	Tail frequency
Backbone RMSD	0.276 Å	0.477 Å	$24.33\% > 0.5$ Å; $8.94\% > 1$ Å
Side-chain RMSD	0.403 Å	0.815 Å	$42.18\% > 0.5$ Å; $24.93\% > 1$ Å
Maximum $\lvert\Delta\chi\rvert$	8.43°	34.93°	$24.62\% > 30^\circ$; $20.93\% > 60^\circ$

A.4 APO–HOLO STERIC RELIEF

For matched protein heavy atom i, ligand heavy atom j, and pocket state $Q \in \{A, H\}$, let $d_{ij}^{Q} = \|\mathbf{x}_i^{Q} - \mathbf{x}_j^{L}\|_2$. Using Bondi van der Waals radii (Bondi, 1964), we define

$$p_{ij}^{Q}(\alpha) = \left[\alpha(r_i^{\mathrm{vdW}} + r_j^{\mathrm{vdW}}) - d_{ij}^{Q}\right]_+, \qquad C_{A|H}(\alpha, \delta) = \mathbf{1}\left[\max_{i,j}\{p_{ij}^{A} - p_{ij}^{H}\} > \delta\right]. \tag{17}$$

The same aligned crystallographic ligand and the same protein–ligand atom pair are used in both states. Thus, a legitimate short contact already present in the holo complex serves as a reference rather than being automatically classified as invalid. Using $\alpha = 0.75$ and $\delta = 0.20$ Å, 44.29% of the 24,500 LMDB complexes exhibit apo-specific steric occlusion.

Under an optimistic attribution in which backbone atoms and C_β (C_α for Gly) remain non-retractable, 76.54% of the occluded complexes involve distal side-chain atoms only. This supports the potential utility of local side-chain relaxation, but does not imply that every case admits a feasible rotamer.

Table 4: Sensitivity of holo-calibrated apo-specific steric occlusion. "Hard/root" is measured over all complexes, whereas "distal only" is measured among occluded complexes.

α	δ (Å)	Occluded	Hard/root	Distal only
0.70	0.20	37.31%	8.02%	78.49%
0.75	0.10	50.85%	12.56%	75.31%
0.75	0.20	44.29%	10.39%	76.54%
0.75	0.30	38.79%	8.73%	77.48%
0.80	0.20	52.41%	13.94%	73.40%

A.5 LEAD-GUIDED SIDE-CHAIN-FLEXIBLE DOCKING

This independent experiment tests whether inexpensive local relaxation can produce useful conditions for a fixed-pocket generator. It does not construct ApoFlowMol targets or inputs. The complete procedure is:

1. infer donor, acceptor, aromatic, and hydrophobic interaction centers from the apo pocket and rank candidate leads using only pre-docking pocket and ligand features;
2. nominate at most ten non-Gly/Ala side chains within 5 Å of the selected center, excluding backbone atoms;
3. use SMINA to optimize ligand pose and torsions jointly with those side chains while fixing the backbone (Trott & Olson, 2010; Koes et al., 2013);
4. run three seeds in a 16 Å cubic box, retaining five modes per seed at exhaustiveness 8;
5. accept a candidate when at least two seeds recover the intended local interaction with no severe clash and at most one soft clash; copy only the accepted side-chain heavy-atom coordinates into the complete apo pocket;
6. condition FLOWR on the selected pocket and evaluate its generated ligands (Cremer et al., 2026).

Writing $I_{cq} = 1$ when candidate c passes the physical checks in seed q, the consensus rule is

$$y_c = \mathbf{1}\left[\sum_{q=1}^{3} I_{cq} \geq 2\right]. \tag{18}$$

A two-layer, 64-unit multilayer perceptron provides rapid preselection from pre-docking pocket and ligand features. It is trained on 1,000 candidates from 50 interaction centers using a center-wise 80/20 split. On ten held-out centers, the ranker attains average precision 0.736, a 90% top-1 positive rate, and 72% precision among the top five candidates per held-out center, compared with 0.511 and 60% top-1 for the original ranking score. Complete library-wise and exhaustive pocket-selection results are reported in Appendix G.4.

A.6 SEQUENCE-DEFINED PEPTIDE-CONTINUITY AUDIT

Expected peptide links are defined from adjacent residue records in the original, uncropped apo chain, retaining chain identifiers, residue numbers, and insertion codes. A link is recognized only when both C_i and N_{i+1} are present and $1.0 < d^A(C_i, N_{i+1}) < 2.0$ Å. Thus pocket cropping, missing residues, or generated spatial proximity cannot create a new expected link. For each expected link we report

$$e_i = |d^{\text{gen}}(C_i, N_{i+1}) - 1.33\,\text{Å}|, \qquad \text{Break}_\tau = \mathbf{1}[e_i > \tau], \quad \tau \in \{0.10, 0.20\}\,\text{Å}. \tag{19}$$

The 0.20 Å threshold is the conservative break criterion; the 0.10 Å rate and continuous errors show sensitivity to the chosen tolerance. Residues are paired by the corrected apo–holo mapping rather than by generated spatial adjacency.

Table 5: Peptide-link errors on 100 matched systems. Each control has 1,393 apo-defined links; each 500-output cohort has 6,965, with no missing expected links. Links pools all link instances; Pockets counts outputs with at least one flagged link. Ave/Med aggregate the per-output C–N MAE with equal output weights (five outputs per system for generated cohorts). Max is the largest individual absolute error, not a bond length. All errors are relative to 1.33 Å; values are in angstroms.

		$e_i > 0.10$ Å		$e_i > 0.20$ Å		Pocket MAE		
Pocket source	n	Links	Pockets	Links	Pockets	Ave	Med	Max
Apo input	100	0.00%	0.00%	0.00%	0.00%	0.007	0.006	0.048
Experimental holo	100	0.00%	0.00%	0.00%	0.00%	0.008	0.007	0.067
Apo2Mol	500	89.56%	99.60%	79.84%	99.40%	0.831	0.757	8.996
YuelDesign	500	28.89%	79.00%	28.89%	79.00%	1.155	0.997	8.781
Link-GR (ODE)	500	0.00%	0.00%	0.00%	0.00%	0.007	0.006	0.048
Link-GR (SDE)	500	0.00%	0.00%	0.00%	0.00%	0.007	0.006	0.049

B CORRESPONDENCE AND REACHABLE TARGETS

Identity-preserving correspondence. A trusted apo–holo residue correspondence requires all mapped atoms to identify one unused holo residue, matching residue identities, and matching paired atom names. We then align matched C_α atoms with Kabsch least squares (Kabsch, 1976). Robustness is supplied by the following deterministic wrapper: initialize from both median translation and an all-candidate Kabsch fit; retain the initialization with the most residuals below 8 Å; refit its inliers for at most three rounds; and apply the final rigid transform to all holo protein and ligand atoms. Thus "robust Kabsch" refers to Kabsch inside an inlier/refit procedure, not to a different closed-form rotation.

With $\mathbf{c}_r$ denoting residue anchors, the trusted correspondence mask is

$$m_r^{\text{match}} = m_r^{\text{identity}} \mathbf{1}[\|\mathbf{R}_g \mathbf{c}_r^H + \mathbf{t}_g - \mathbf{c}_r^A\|_2 \le 8\,\text{Å}]. \tag{20}$$

Untrusted residues remain visible as apo context. Correspondence and reachability masks restrict protein-motion, protein-coordinate, motion-recovery, and local/cross-distance supervision to valid targets. Pharmacophore and clash losses instead use structural atom masks, so these context atoms may still contribute to those terms.

In the training cache, identity-preserving correspondence covers 928,775 of 932,028 residues (99.6510%). After the additional trust and reachability checks, 924,903 residues (99.2355%) admit a reachable linked coordinate target. We define an interaction-core residue as one having a heavy atom within 4.5 Å of the ligand in apo or aligned holo coordinates. Of 319,499 such residues, 2,928 (0.9164%) lack reachable coordinate supervision but remain in the structural context.

Symmetry and reachable coordinates. For a chemical torsion with physical period $p_{r,k}$,

$$\Delta\chi^*_{r,k} = \text{wrap}_{p_{r,k}}(\chi^H_{r,k} - \chi^A_{r,k}), \tag{21}$$

where $p = \pi$ for ASP χ_2, GLU χ_3, PHE χ_2, and TYR χ_2, and $p = 2\pi$ otherwise. Symmetry-equivalent atom names are canonicalized before coordinate supervision; Pro side-chain torsions are

frozen because tree kinematics does not close its ring. For Link, protein coordinates are supervised against $\widetilde{\mathcal{P}}^{H}_{\mathrm{Link}} = K_{\mathrm{Link}}(\mathcal{P}^{A}, \mathcal{S}_1^{\mathrm{Link},*})$; the fixed-backbone χ coordinate auxiliary instead uses aligned holo atoms. Link target fitting follows the same kinematic order and constraints as the sampler, preventing optimization toward an unreachable Cartesian target.

Reachable-state fitting. Fit_m is deterministic and uses no iterative optimizer. For Link, each segment-root translation is the matched C_α displacement, and its rotation is obtained by a Kabsch fit of apo and holo N/C vectors, including O when available, about the root C_α. We then extract admissible backbone-angle changes and wrapped $\phi/\psi/\omega$ differences, reconstruct each segment from root to tail, and fit residual χ_0 and $\chi_{1:4}$ updates against that reconstruction. Targets whose upstream linked path is untrusted are masked. For the χ representation, the apo N–C_α–C coordinates are retained and only the wrapped carbonyl and side-chain torsion targets are extracted.

C Geometry-Constrained Protein States

Fixed-backbone χ state. For pivot $\mathbf{p}$, unit bond axis $\mathbf{u}$, torsion increment δ, and downstream atom $\mathbf{x}$, the update is the Rodrigues rotation

$$\mathbf{x}' = \mathbf{p} + \mathbf{R}(\mathbf{u}, \delta)(\mathbf{x} - \mathbf{p}). \tag{22}$$

ApoFlowMol-χ rotates carbonyl O/OXT about C_α–C through χ_0, then applies residue-dependent $\chi_{1:4}$ from proximal to distal. Only descendants of each chemical bond move; the apo N–C_α–C trace is fixed.

Chain-linked state. ApoFlowMol-Link assigns one root transform $(\mathbf{t}_s, \mathbf{R}_s)$ to each recognized continuous peptide segment. From root to tail, it composes linked $\phi/\psi/\omega$ torsions and the three backbone angles C_{i-1}–N_i–$C_{\alpha i}$, N_i–$C_{\alpha i}$–C_i, and $C_{\alpha i}$–C_i–N_{i+1}; carbonyl χ_0 and side-chain $\chi_{1:4}$ follow. Every joint moves its complete downstream subtree. Backbone-angle updates use configuration-specific bounds and optional learned gates. For Link targets, χ_0 is fitted after linked-backbone reconstruction as the wrapped residual carbonyl rotation. It is retained only when the residue-alignment residual is at most 0.5 Å and $|\Delta\chi_0| \le 30^\circ$. This degree of freedom preserves the carbonyl C–O bond length but does not by itself constrain O_i–C_i–N_{i+1} peptide-plane geometry.

For Link-GR and Link-R, let $\mathcal{A}^{A}_{y,a}$ contain angle type a for amino-acid type y in training apo structures. We set

$$\begin{aligned} l_{y,a} &= Q_{0.005}(\mathcal{A}^{A}_{y,a}) - 1^\circ, & u_{y,a} &= Q_{0.995}(\mathcal{A}^{A}_{y,a}) + 1^\circ, \\ \widetilde{l}_{r,a} &= \min(l_{y_r,a}, \alpha^{A}_{r,a}), & \widetilde{u}_{r,a} &= \max(u_{y_r,a}, \alpha^{A}_{r,a}), \\ \Pi_{r,a}(\Delta\alpha) &= \mathrm{clip}(\alpha^{A}_{r,a} + \Delta\alpha, \widetilde{l}_{r,a}, \widetilde{u}_{r,a}) - \alpha^{A}_{r,a}. \end{aligned} \tag{23}$$

Thus an apo outlier is not forced discontinuously into the population interval, but cannot move farther outward. Holo targets outside the effective interval are masked rather than clipped. The retained trusted-angle fractions are 99.6%, 99.8%, and 99.9% in train, validation, and test, respectively.

Backbone-angle variants. Link-G5 uses a learned gate with a cumulative apo-relative $\pm 5^\circ$ bound; Link-GR combines the gate with Eq. 23; Link-R uses that range without gating. For gated variants, the training target is

$$y_{r,a} = \mathrm{clip}\left(\frac{|\Delta\alpha^{*}_{r,a}| - 1^\circ}{0.5^\circ}, 0, 1\right),$$

and the gate is trained by binary cross-entropy. At each sampling step, the predicted clean angle is accepted when $\mathrm{sigmoid}(h_{r,a}) \ge 0.5$; otherwise the current cumulative angle is retained before range projection. Final cumulative changes below 1° are set to zero. Link-R omits the gate head and loss.

The root transform is rigid; each internal-coordinate update preserves the two adjacent bond lengths; and every later update moves both bond endpoints rigidly or rotates around that bond. Induction along the segment therefore preserves apo-template N–C_α, C_α–C, and recognized C_i–N_{i+1} distances at every integration step, without post-hoc peptide projection.

D Co-Flow and Predictor Details

D.1 Continuous and categorical paths

Ligand Gaussian points are permuted by one Hungarian assignment and rotated about the apo-pocket center by one Kabsch solve; no target-dependent translation is introduced (Klein et al., 2023; Cremer et al., 2026). For an unconstrained Euclidean modality u, the SDE training bridge is

$$z_t^u = (1-t)z_0^u + tz_1^u + \gamma_u(t)\boldsymbol{\epsilon}, \quad \gamma_u(t) = \sigma_u\sqrt{t(1-t)}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{24}$$

Its conditional pathwise velocity and score are $b_t^{u,*} = z_1^u - z_0^u + \dot{\gamma}_u(t)\boldsymbol{\epsilon}$ and $s_t^{u,*} = -\boldsymbol{\epsilon}/\gamma_u(t)$. At the population optimum of squared endpoint regression, the endpoint field and the corresponding marginal velocity and score satisfy $q^u = b^u + \frac{1}{2}\sigma_u^2 s^u$, motivating the sampling update

$$dz_t^u = \left[q_\theta^u + \tfrac{1}{2}(g_u^2 - \sigma_u^2)s_\theta^u\right] dt + g_u\, dW_t^u, \qquad g_u^2 = 2\sigma_u^2(1-t). \tag{25}$$

Periodic states use wrapped displacements and wrapped-normal score targets. Rotations use isotropic tangent noise composed on $\mathrm{SO}(3)$ and a tangent-normal score approximation; backbone-angle states use the cumulative $\pm 5^\circ$ bound for Link-G5 and Eq. 23 for Link-GR and Link-R. We apply the same drift correction to these states without claiming exact marginal preservation after wrapping, manifold approximation, or projection. ODE training omits bridge noise, and ODE sampling uses $dz_t^u = q_\theta^u\, dt$. The noise scales are 2.0 Å for ligand coordinates and segment translation, 0.005 rad for segment rotation, 1.0 rad for torsions, and 1° for backbone angles; the first four modality choices follow DynamicFlow (Zhou et al., 2025). Training uses antithetic times clipped to $[10^{-3}, 1-10^{-3}]$; denominators of the form $1-t$ are clamped from below at 10^{-3}. Sampling evaluates the integration field at $t_k = k/56$, $k = 0, \ldots, 55$. The additional evaluation at $t = 1$ is used only for final categorical decoding and endpoint projection, not for another continuous integration step.

Atom and bond categories follow the main-text categorical bridge. A non-absorbing correction process permits later revision of an earlier class. Bond updates are sampled on the strict upper triangle and mirrored, ensuring a symmetric zero-diagonal bond matrix. During training, a detached preliminary pass supplies atom/bond endpoint probabilities with probability 0.5; during sampling, the previous-step probabilities provide the same soft self-conditioning. For current class i, endpoint probabilities p_θ, class count K, and step size Δt, the implemented off-diagonal transition is

$$\widetilde{P}_{ij} = \mathrm{clip}\left(\Delta t\left[\frac{1+2\eta(K-1)t}{1-t}p_\theta(j) + \eta p_\theta(i)\right], 0, 1\right), \qquad j \neq i, \tag{26}$$

where $\eta = 1$ except for the final transition, where $\eta = 0$. We set $\widetilde{P}_{ii} = [1 - \sum_{j\neq i}\widetilde{P}_{ij}]_+$, renormalize the row, and sample the next class. Bond transitions are sampled on the strict upper triangle and mirrored.

D.2 Hierarchical equivariant predictor

At every step, the pocket is reconstructed before rebuilding the complex graph. Periodic internal states enter protein atom features through sine–cosine embeddings, and backbone-angle states through normalized scalar embeddings; segment-root motion enters through reconstructed coordinates. Atom-to-residue pooling returns invariant residue features and separate equivariant vector channels. The residue Transformer receives only invariant features, C_α distances, and chain-link biases, while the vector channels bypass it and parameterize equivariant segment-root heads.

Rigid-motion convention. Under a global transform $\mathbf{x}' = Q\mathbf{x} + \mathbf{b}$, scalar features and internal angles are invariant, while pooled vector channels satisfy $\mathbf{v}' = Q\mathbf{v}$. Segment translations and global-axis rotation vectors transform as $\mathbf{t}_s' = Q\mathbf{t}_s$ and $\mathbf{r}_s' = Q\mathbf{r}_s$, giving $\mathbf{R}_s' = Q\mathbf{R}_s Q^\top$. Consequently, $\widehat{\mathbf{X}}_1' = Q\widehat{\mathbf{X}}_1 + \mathbf{b}$. No Cartesian vector component is concatenated to the invariant residue Transformer input. Random global rotation/translation regression tests are used to verify these transformations numerically.

Following the interaction-semantics idea of DiffPharma, we derive donor, acceptor, hydrophobe, aromatic, cation, and anion sites solely from apo atom and residue identities (Sako et al., 2026).

Table 6: Shared ApoFlowMol predictor architecture used across all configurations.

Component	Function
Full-atom complex graph	Ligand and protein atoms; ligand, cross-complex, intra-residue, recognized peptide-link, and geometric-neighbor edges.
Atom-level encoder	Sparse $\mathrm{SE}(3)$-equivariant UniTransformer, hidden size 128, nine layers, 16 heads, $k = 32$; graph neighbors refresh every three layers.
Ligand heads	Clean coordinate endpoint, atom classes, symmetric bond classes, and symmetric bond-length endpoint.
Residue network	Attention pooling from protein atoms followed by a three-layer geometry-aware residue Transformer.
Protein conditioning	Invariant ligand/categorical context and internal-state embeddings; reconstructed geometry and separate equivariant vector channels.
Protein heads	Either carbonyl/side-chain χ endpoints or the Link state defined in Appendix C; Link-G5 and Link-GR additionally predict a backbone-angle gate.
Ligand size	Independent equivariant encoder using only the static apo pocket.

Their coordinates move with the reconstructed pocket. Ligand–site compatibility attention is recomputed at each step; neither holo coordinates nor the reference ligand are used to form sites at inference.

E Objectives and Optimization

Loss reductions are term-specific: coordinate, joint, ligand-geometry, pharmacophore, and score losses average valid elements across the minibatch, whereas local-distance, cross-distance, and clash losses average per-complex quantities. Size and atom classes use unweighted cross-entropy. Bond classes use weighted cross-entropy over valid upper-triangular pairs, with weights $(0.12, 1.0, 1.5, 2.0, 1.3)$ for no bond, single, double, triple, and aromatic bonds, normalized by the sum of target-class weights. Euclidean endpoints use MSE or smooth-L_1; torsions use wrapped squared errors; and rotations use squared $\mathrm{SO}(3)$ geodesic errors. With m denoting the selected protein representation,

$$\begin{aligned}
\mathcal{L} &= \mathcal{L}_X + 0.3\mathcal{L}_A + 0.3\mathcal{L}_E + 0.15\mathcal{L}_N + \mathcal{L}^m_{\text{prot}} \\
&\quad + 0.3\mathcal{L}_{\text{bondlen}} + 0.3\mathcal{L}_{\text{lenend}} + 0.1\mathcal{L}_{\text{bondangle}} + 0.1\mathcal{L}_{\text{motion}} \\
&\quad + \mathbf{1}[t > 0.65](0.5\mathcal{L}_{\text{int}} + 0.75\mathcal{L}_{\text{cross}} + 0.30\mathcal{L}_{\text{pharm}} + 0.30\mathcal{L}_{\text{clash}}) + 0.5\mathbf{1}_{\text{SDE}}\mathcal{L}_{\text{score}}, \\
\mathcal{L}^{\chi}_{\text{prot}} &= 0.5\mathcal{L}_{\chi} + 1.0\mathcal{L}_{\chi_0} + 0.5\mathcal{L}^P_{\text{coord}}, \\
\mathcal{L}^{\text{Link}}_{\text{prot}} &= 0.5(\mathcal{L}_{tr} + \mathcal{L}_{rot} + \mathcal{L}_{\chi} + \mathcal{L}_{bb} + \mathcal{L}_{\angle} + \mathcal{L}^P_{\text{coord}}) + 1.0\mathcal{L}_{\chi_0} \\
&\quad + 0.2\mathbf{1}_{\{\text{Link-G5,Link-GR}\}}\mathcal{L}_{\text{gate}}.
\end{aligned} \tag{27}$$

Peptide geometry is logged as a diagnostic outside the optimized objective: it is identically zero for Link reconstruction and constant with respect to the trainable parameters of the fixed-backbone model. Geometry losses supervise reference bond distances, the predicted bond-length endpoint, and bonded angles. $\mathcal{L}_{\text{int}}$ covers local complex distances; $\mathcal{L}_{\text{cross}}$ isolates protein–ligand pairs; and $\mathcal{L}_{\text{pharm}}$ supervises compatible typed sites. Following DynamicFlow, only these fine interaction terms and clash supervision begin after $t > 0.65$; ligand and protein flow losses remain active for all t (Zhou et al., 2025).

For ligand atom i and protein atom j, let r_{ij} be their summed van der Waals radius, d_{ij} the predicted distance, and d^*_{ij} the matched reference distance. We use

$$\begin{aligned}
c^{\text{rel}}_{ij} &= \left[[0.8r_{ij} - d_{ij}]^2_+ - [0.8r_{ij} - d^*_{ij}]^2_+\right]_+, \\
c^{\text{abs}}_{ij} &= [\min(0.7r_{ij}, d^*_{ij}) - d_{ij}]^2_+, \qquad \mathcal{L}_{\text{clash}} = \text{mean}_i \max_j (c^{\text{rel}}_{ij} + c^{\text{abs}}_{ij}) + 0.5 \max_{i,j} (c^{\text{rel}}_{ij} + c^{\text{abs}}_{ij}).
\end{aligned} \tag{28}$$

This preserves experimentally close contacts while preventing a single severe collision from being diluted by many distant protein atoms. The reference array used for d^*_{ij} is defined for every

structurally masked atom and is independent of the coordinate-supervision mask. For Link it is the reachable linked reconstruction; for the χ model it is the aligned holo array, with apo-template coordinates used when a matched holo atom is absent. Thus, context atoms can enter clash supervision even when they are excluded from coordinate supervision.

Models use AdamW with learning rate 2×10^{-4}, $\beta = (0.95, 0.999)$, weight decay 10^{-4}, 1,000 warmup updates followed by cosine decay, mixed precision, effective batch size 16, and gradient clipping at 8.0. Training runs for at most 200,000 optimizer updates, validates on all 1,071 validation systems each epoch, and stops after ten epochs without an EMA-loss improvement of 10^{-4}. Raw parameters and an exponential moving average (decay 0.998) are evaluated and selected independently; model snapshots are retained every five epochs.

F Sampling, Refinement, and Evaluation

Joint sampling. The sampler uses 56 uniform steps and updates ligand and pocket at every step:

1. reconstruct $\mathcal{P}_t = K_m(\mathcal{P}^A, \mathcal{S}_t^m)$ and build the current full-atom complex graph;
2. predict ligand endpoints, the selected protein endpoint, and continuous-state scores for SDE sampling;
3. for ApoFlowMol-Link, apply bounded ligand bond-length corrections to the predicted clean endpoint before constructing its integration field; advance all active continuous states from the same pre-step prediction, apply the configured backbone-angle gate when present, and project cumulative angle changes onto the admissible interval;
4. update atom and bond categories, mirror bond updates, and store soft self-conditioning;
5. reconstruct the updated pocket and re-encode the complex at the next step.

ApoFlowMol-χ applies the predicted bond-length correction at final decoding rather than within each integration step. After the 56 updates, both variants perform a final network evaluation for categorical decoding and ligand-only bond-length projection. These bond corrections never move protein atoms. The output is one ligand and one kinematically reconstructed full-atom pocket.

Refinement for pocket conditioning. The pocket-conditioning protocol enables joint physical guidance after $t = 0.65$ and final force-field refinement before candidate acceptance. Guidance optimizes ligand internal geometry and ligand–current-pocket clashes within a trust region; protein responses are restricted to Link variables. A final MMFF94s ligand minimization, with UFF fallback, limits each heavy-atom displacement to 0.15 Å, after which at most two bounded Link updates allow the pocket to follow. No holo structure, reference ligand, or docking score enters this refinement.

Evaluation. Ligand metrics include sanitization validity, single-component connectivity, QED, synthetic accessibility, Lipinski properties, bond-length error, intramolecular clash, protein clash, interaction recovery, and Vina-based scores. Pocket metrics include apo-to-generated, generated-to-holo, and generated-to-reachable-target RMSD, together with the sequence-defined peptide-link errors of Eq. 19.

For the pocket-conditioning benchmark, we retain five ApoFlowMol ligand–pocket candidates per test system after refinement and candidate checks.

Vina-GPU docking scores do not guide the neural generation trajectories. After generation, each accepted candidate is docked into its corresponding generated pocket using Vina-GPU-2.1, with a 20 Å cubic box and search depth 1. We select the pocket associated with the lowest docking score and condition FLOWR on that pocket to obtain 100 valid, connected, reference-size ligands with unique canonical SMILES. These ligands are evaluated using the same Vina-GPU protocol. The accepted ApoFlowMol candidates and downstream FLOWR ligands form separate evaluation sets.

For a stated candidate budget K, the best-score summary is

$$V_s^{\mathrm{best},K} = \min_{k \in \mathcal{C}_s^{(K)}} V_{s,k}, \qquad \overline{V}^{\mathrm{best},K} = \frac{1}{S} \sum_{s=1}^{S} V_s^{\mathrm{best},K},$$

where $\mathcal{C}_s^{(K)}$ is the successfully scored candidate set for test pocket s, and S is the number of evaluated pockets. Best-of-5 and best-of-100 results are distinguished.

G Additional Experimental Details and Results

G.1 Evaluation protocol

All reported direct-generation and pocket-reuse metrics use the same sequence-diverse 100-system subset described in Appendix A.2. For ApoFlowMol, five ligand–pocket candidates are retained per system after the sampling and refinement procedure in Appendix F; the candidate with the lowest successful Vina-GPU score defines the direct result and the pocket used in the reuse experiment. Apo2Mol and YuelDesign are evaluated from their corresponding generated ligand–pocket outputs.

For pocket reuse, the selected apo, holo, or generated pocket is held fixed while FLOWR produces reference-size ligands. FLOWR-5 and FLOWR-100 select the lowest successfully docked ligand among the first 5 and 100 saved outputs, respectively. Vina-GPU 2.1 uses a 20 Å cubic box and search depth 1. Docking is applied only after generation and does not guide the ApoFlowMol or FLOWR trajectories. Every reported condition covers all 100 systems.

QED, normalized SA, LogP, and Lipinski counts are computed for the ligand selected by Vina. We use $\mathrm{SA}_{\mathrm{norm}} = (10 - \mathrm{SA}_{1:10})/9$. Lipinski$_5$ counts exact molecular mass < 500, hydrogen-bond donors ≤ 5, acceptors ≤ 10, and $-2 \leq \log P \leq 5$. HA is the percentage scoring strictly below the matched reference ligand in its experimental holo pocket.

G.2 Link backbone-angle variants

The main text uses *ApoFlowMol-Link* for the primary Link-GR parameterization. We additionally evaluate two backbone-angle controls: Link-G5 uses a learned gate with a cumulative apo-relative $\pm 5^\circ$ bound, Link-GR combines the gate with amino-acid-specific empirical angle ranges, and Link-R retains the empirical ranges without gating. All three preserve the same chain-linked reconstruction, segment-root motion, linked $\phi/\psi/\omega$ torsions, carbonyl χ_0, and side-chain $\chi_{1:4}$ updates described in Appendix C.

Table 7: Backbone-angle mechanisms used by ApoFlowMol-Link. Empirical ranges are estimated only from training apo structures.

Configuration	Learned gate	Angle constraint	Final dead zone
Link-G5	Yes	$\lvert\Delta\alpha\rvert \leq 5^\circ$	1°
Link-GR	Yes	Empirical absolute range	1°
Link-R	No	Empirical absolute range	None

Table 8: Direct generation and pocket-reuse results for the six ApoFlowMol-Link configurations. Vina entries are Avg./Med. in $\mathrm{kcal\,mol^{-1}}$; HA is shown in parentheses.

Configuration	Direct	FLOWR-5	FLOWR-100
Link-G5 (ODE)	$-7.70/-7.60$ (76%)	$-7.17/-7.00$ (49%)	$-8.53/-8.35$ (88%)
Link-G5 (SDE)	$-7.73/-8.00$ (71%)	$-7.60/-7.75$ (67%)	$-8.75/-8.90$ (89%)
Link-GR (ODE)	$-7.73/-7.65$ (77%)	$-7.18/-6.95$ (51%)	$-8.52/-8.45$ (89%)
Link-GR (SDE)	$-7.90/-8.10$ (79%)	$-7.68/-7.60$ (68%)	$-8.82/-8.80$ (90%)
Link-R (ODE)	$-7.81/-7.75$ (79%)	$-7.21/-7.10$ (53%)	$-8.65/-8.60$ (86%)
Link-R (SDE)	$-7.48/-7.90$ (66%)	$-7.61/-7.60$ (61%)	$-8.80/-8.80$ (86%)

For concise main-text presentation, Link-GR was selected after exploratory comparison of the three Link variants because its SDE model gives the lowest direct and FLOWR-100 mean Vina scores while retaining both a learned activation gate and empirical angle ranges. Comparisons among Link-G5, Link-GR, and Link-R are therefore interpreted descriptively rather than as an independently validated model-selection result. Its ODE variant gives higher direct QED and normalized

SA than its SDE variant. Across all three angle mechanisms, the reconstructed peptide links retain approximately 0.006 Å median C–N MAE.

Paired comparisons. For each comparison, we compute the per-system Vina-score difference as the first configuration minus the second and estimate its mean and 95% confidence interval using 10,000 paired bootstrap resamples. Negative values favor the first configuration.

Table 9: Paired Vina-score differences between Link configurations, in $\mathrm{kcal\,mol^{-1}}$.

Comparison	Output	Difference	95% CI
Link-G5 SDE − ODE	Direct	-0.03	[-0.27,0.20]
Link-G5 SDE − ODE	FLOWR-100	-0.22	[-0.51,0.07]
Link-GR SDE − ODE	Direct	-0.17	[-0.43,0.10]
Link-GR SDE − ODE	FLOWR-100	-0.31	[-0.63,0.03]
Link-R SDE − ODE	Direct	0.34	[0.07,0.62]
Link-R SDE − ODE	FLOWR-100	-0.15	[-0.42,0.13]
Link-GR ODE − Link-G5 ODE	Direct	-0.03	[-0.21,0.15]
Link-GR ODE − Link-G5 ODE	FLOWR-100	0.02	[-0.18,0.22]
Link-R ODE − Link-GR ODE	Direct	-0.08	[-0.24,0.08]
Link-R ODE − Link-GR ODE	FLOWR-100	-0.13	[-0.33,0.07]
Link-GR SDE − Link-G5 SDE	Direct	-0.17	[-0.41,0.08]
Link-GR SDE − Link-G5 SDE	FLOWR-100	-0.07	[-0.39,0.27]
Link-R SDE − Link-GR SDE	Direct	0.42	[0.20,0.64]
Link-R SDE − Link-GR SDE	FLOWR-100	0.03	[-0.22,0.27]

G.3 Expanded ligand-property and pocket-conditioning results

Tables 10 and 11 report all eight ApoFlowMol configurations. The main text uses Link-GR as ApoFlowMol-Link and reports the two dynamics separately.

Table 10: Direct-generation results for all eight ApoFlowMol configurations. Each entry is Avg./Med. over one selected ligand per system; Vina is in $\mathrm{kcal\,mol^{-1}}$.

Configuration	Vina ↓	QED ↑	SA ↑	LogP	Lipinski ↑	HA (%) ↑
χ (ODE)	$-6.97/-7.15$	$0.43/0.42$	$0.53/0.53$	$0.52/0.43$	$4.55/5.00$	57
χ (SDE)	$-7.25/-7.15$	$0.46/0.42$	$0.48/0.47$	$0.49/0.33$	$4.56/5.00$	65
Link-G5 (ODE)	$-7.70/-7.60$	$0.46/0.45$	$0.47/0.45$	$0.42/0.41$	$4.60/5.00$	76
Link-G5 (SDE)	$-7.73/-8.00$	$0.40/0.38$	$0.39/0.37$	$-1.06/-1.20$	$4.33/5.00$	71
Link-GR (ODE)	$-7.73/-7.65$	$0.48/0.47$	$0.48/0.47$	$0.28/0.57$	$4.59/5.00$	77
Link-GR (SDE)	$-7.90/-8.10$	$0.39/0.37$	$0.32/0.28$	$-1.51/-1.52$	$4.35/5.00$	79
Link-R (ODE)	$-7.81/-7.75$	$0.46/0.46$	$0.48/0.49$	$0.73/0.48$	$4.62/5.00$	79
Link-R (SDE)	$-7.48/-7.90$	$0.40/0.36$	$0.41/0.39$	$-0.57/-0.69$	$4.48/5.00$	66

G.4 FlexDock library and exhaustive pocket search

This experiment tests whether lead-guided local adjustment of apo pockets can provide more effective receptor inputs for fixed-pocket molecular generation. Every proposal starts from the same apo structure; only selected side chains near typed interaction centers are allowed to move, while the backbone and the remainder of the pocket stay fixed. We evaluate the adapted pockets through downstream FLOWR generation and docking, retaining the generator's original parameters and fixed-pocket sampling procedure. The library-wise and exhaustive analyses below assess how the resulting performance changes with pocket selection breadth.

Each lead library provides single-center proposals and all geometrically compatible combinations of two or more interaction centers. The two-stage protocol generates five FLOWR ligands for every candidate pocket, selects one pocket per system and library by its best Vina-GPU score, and extends only that pocket to 100 ligands. The exhaustive protocol instead evaluates 100 FLOWR ligands for every candidate pocket. The five-library rows additionally select the best result across libraries for each system. Across the five libraries, the candidate set contains 1,205 single-center and 1,194 multi-center pockets.

Table 11: Expanded pocket-reuse results for all eight ApoFlowMol configurations. The selected pocket is fixed during FLOWR generation. Entries are Avg./Med.; Vina is in $\mathrm{kcal\,mol^{-1}}$.

Configuration	Selection	Vina ↓	QED ↑	SA ↑	LogP	Lipinski ↑	HA (%) ↑
χ (ODE)	FLOWR-5	$-6.99/-6.70$	$0.56/0.56$	$0.72/0.72$	$1.97/2.28$	$4.74/5.00$	48
χ (ODE)	FLOWR-100	$-8.28/-7.95$	$0.60/0.61$	$0.68/0.68$	$2.11/2.04$	$4.76/5.00$	81
χ (SDE)	FLOWR-5	$-6.87/-6.55$	$0.57/0.55$	$0.71/0.72$	$1.77/1.84$	$4.75/5.00$	51
χ (SDE)	FLOWR-100	$-8.24/-8.05$	$0.59/0.61$	$0.69/0.69$	$2.36/2.79$	$4.76/5.00$	81
Link-G5 (ODE)	FLOWR-5	$-7.17/-7.00$	$0.57/0.58$	$0.72/0.72$	$1.79/2.36$	$4.71/5.00$	49
Link-G5 (ODE)	FLOWR-100	$-8.53/-8.35$	$0.60/0.61$	$0.70/0.72$	$2.31/2.79$	$4.79/5.00$	88
Link-G5 (SDE)	FLOWR-5	$-7.60/-7.75$	$0.59/0.62$	$0.71/0.74$	$1.78/2.20$	$4.75/5.00$	67
Link-G5 (SDE)	FLOWR-100	$-8.75/-8.90$	$0.61/0.66$	$0.69/0.69$	$2.18/2.54$	$4.72/5.00$	89
Link-GR (ODE)	FLOWR-5	$-7.18/-6.95$	$0.56/0.57$	$0.71/0.70$	$1.91/2.52$	$4.70/5.00$	51
Link-GR (ODE)	FLOWR-100	$-8.52/-8.45$	$0.58/0.58$	$0.68/0.69$	$2.26/2.53$	$4.77/5.00$	89
Link-GR (SDE)	FLOWR-5	$-7.68/-7.60$	$0.56/0.57$	$0.68/0.69$	$1.57/1.74$	$4.65/5.00$	68
Link-GR (SDE)	FLOWR-100	$-8.82/-8.80$	$0.59/0.61$	$0.69/0.70$	$2.12/2.15$	$4.71/5.00$	90
Link-R (ODE)	FLOWR-5	$-7.21/-7.10$	$0.58/0.57$	$0.69/0.69$	$1.91/1.91$	$4.78/5.00$	53
Link-R (ODE)	FLOWR-100	$-8.65/-8.60$	$0.59/0.59$	$0.69/0.69$	$2.30/2.84$	$4.75/5.00$	86
Link-R (SDE)	FLOWR-5	$-7.61/-7.60$	$0.56/0.54$	$0.69/0.70$	$1.68/2.11$	$4.69/5.00$	61
Link-R (SDE)	FLOWR-100	$-8.80/-8.80$	$0.58/0.58$	$0.68/0.68$	$1.99/2.46$	$4.68/5.00$	86

Table 12: **FlexDock library-wise and exhaustive FLOWR-100 results.** Scores are Avg./Med. Vina-GPU in $\mathrm{kcal\,mol^{-1}}$; HA is the percentage below the matched reference ligand in its holo pocket. "Multi" counts systems whose exhaustive winner combines multiple interaction centers.

		Two-stage		Exhaustive		
Lead library	Pockets	Avg./Med.	HA (%)	Avg./Med.	HA (%)	Multi
Enamine860	506	$-8.52/-8.30$	85	$-8.89/-8.75$	89	21
HighFid1920	433	$-8.55/-8.20$	88	$-8.85/-8.65$	90	21
PPI3600	513	$-8.45/-8.30$	84	$-8.83/-8.85$	88	33
Shape1200	488	$-8.52/-8.20$	82	$-8.81/-8.65$	84	26
SinglePharm1500	459	$-8.40/-8.20$	83	$-8.74/-8.60$	85	23
All five	2399	$-9.16/-9.00$	93	$-9.44/-9.30$	96	32

HighFid1920 provides the strongest individual-library two-stage result and is therefore the sole FlexDock row in the main table. Exhaustive search improves the five-library mean Vina-GPU score from -9.16 to -9.44 and HA from 93% to 96%, while increasing the evaluated ligand budget from 59,495 to 239,900. Multi-center pockets account for 32 of the 100 exhaustive winners.

G.5 Peptide geometry across ApoFlowMol configurations

All Link configurations use the same chain reconstruction and preserve the recognized apo-defined peptide links. Figure 13 extends the main-text geometry comparison to all eight ApoFlowMol configurations.

G.6 Ligand-distance and pocket-motion distributions

We compare all eight ApoFlowMol configurations, Apo2Mol, and YuelDesign on the same 100 systems. All-pair and pocket-motion analyses use five outputs per system. Bonded-distance analyses use the explicit molecular graphs and pre-docking coordinates supplied to scoring; no bonds are inferred from distance thresholds and no coordinates are changed.

Distances and JSD. All C–C pairs are unordered pairs of carbon atoms, regardless of bonding. The short-pair view conditions on $d < 2\,\text{Å}$; bonded pairs use explicit SDF edges. Distances are pooled across outputs within each system, normalized per system, and averaged with equal system weights. Experimental references use the same eligible systems. Full pairs use 0–20 Å bins of width $0.10\,\text{Å}$ plus overflow; short pairs use 0–2 Å bins of width $0.02\,\text{Å}$; bonded pairs use 0–3 Å bins of width $0.02\,\text{Å}$ plus overflow.

$$\mathrm{JSD}_2(P,Q) = \tfrac{1}{2}\,\mathrm{KL}_2(P\|M) + \tfrac{1}{2}\,\mathrm{KL}_2(Q\|M), \qquad M = \tfrac{1}{2}(P+Q). \tag{29}$$

We use base-two logarithms, zero contribution for zero-mass terms, and no pseudocounts. Reported values are divergences, not square roots, rounded to three decimals only after computation. Molecules without a C–C bond contribute no bonded distances.

Smooth display, unchanged statistical estimand. All JSDs and numerical summaries use the original unsmoothed data. For display only, common Gaussian bandwidths within each diagnostic are 0.15 Å for full pairs, 0.025 Å for short and bonded pairs, and 0.075 Å for pocket RMSD. Density smoothing reflects at the distance-domain boundaries and preserves the visible histogram mass. Overflow remains in normalization and JSD. Smoothing is used only for readability and does not enter any reported statistic.

Table 13: Ligand C–C pair JSD for the complete five-output cohorts. Short pairs are all pairs below 2 Å and are not restricted to bonds.

Method	All-pair JSD ↓	Short-pair JSD ↓
χ (ODE)	0.028	0.070
χ (SDE)	0.038	0.094
Link-G5 (ODE)	0.031	0.116
Link-G5 (SDE)	0.052	0.134
Link-GR (ODE)	0.030	0.125
Link-GR (SDE)	0.072	0.175
Link-R (ODE)	0.029	0.120
Link-R (SDE)	0.060	0.158
Apo2Mol	0.028	0.157
YuelDesign	0.032	0.166

Common-mask motion. For each system, heavy atoms are intersected across apo, holo, and every output of all ten methods. Seven residue substitutions are excluded. Holo is aligned to apo once using matched C_α atoms; generated pockets stay in their apo frame without an additional fit. The common apo–holo mean/median RMSD is $0.81/0.67$ Å. Motion JSD compares apo–generated and apo–holo RMSDs, using bins of width 0.05 Å over 0–3 Å plus overflow. Each generated method contributes five equally weighted outputs per system.

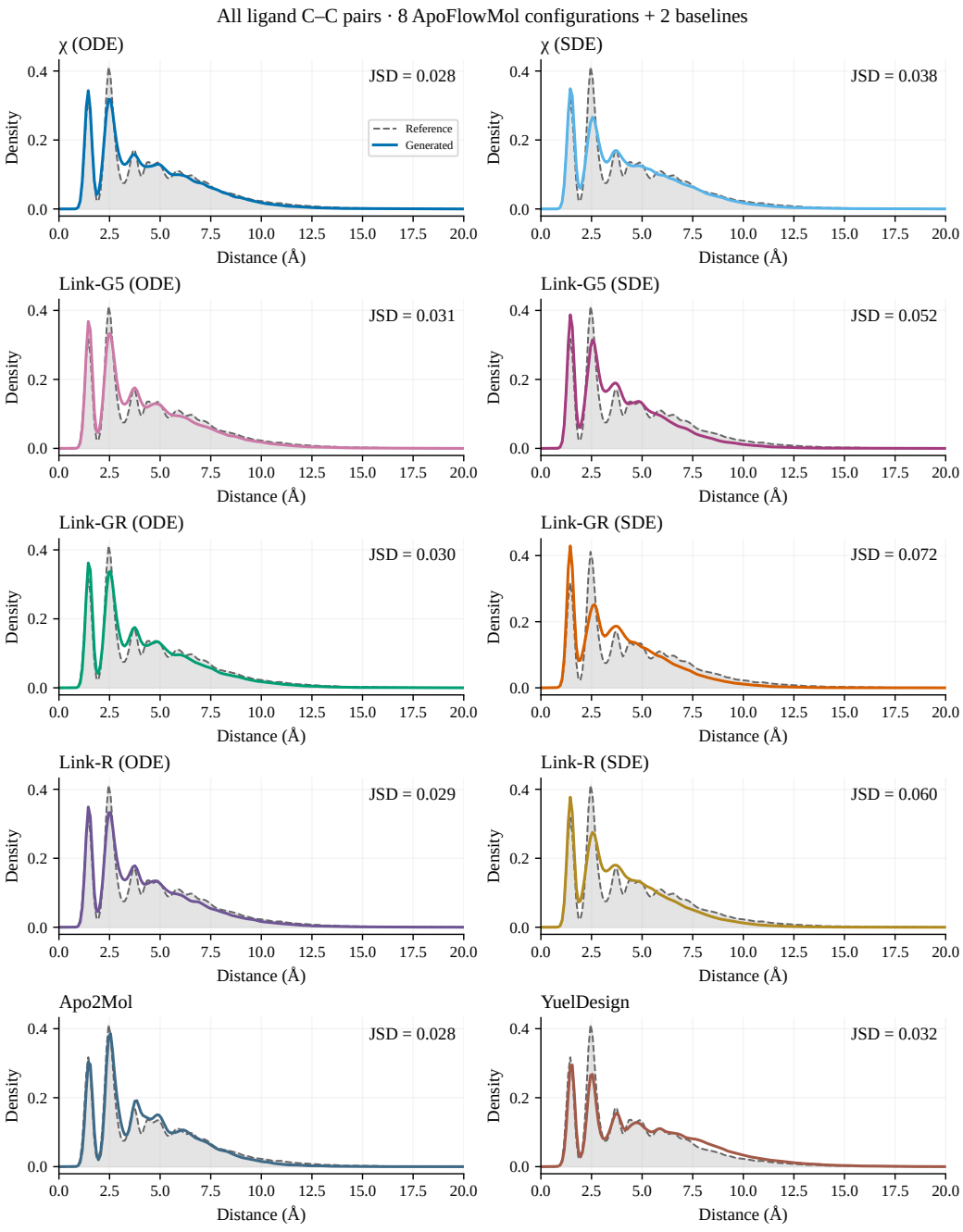


Figure 5: Full ligand C–C pair distributions. Gray denotes the matched reference; JSD uses unsmoothed, system-balanced histograms.

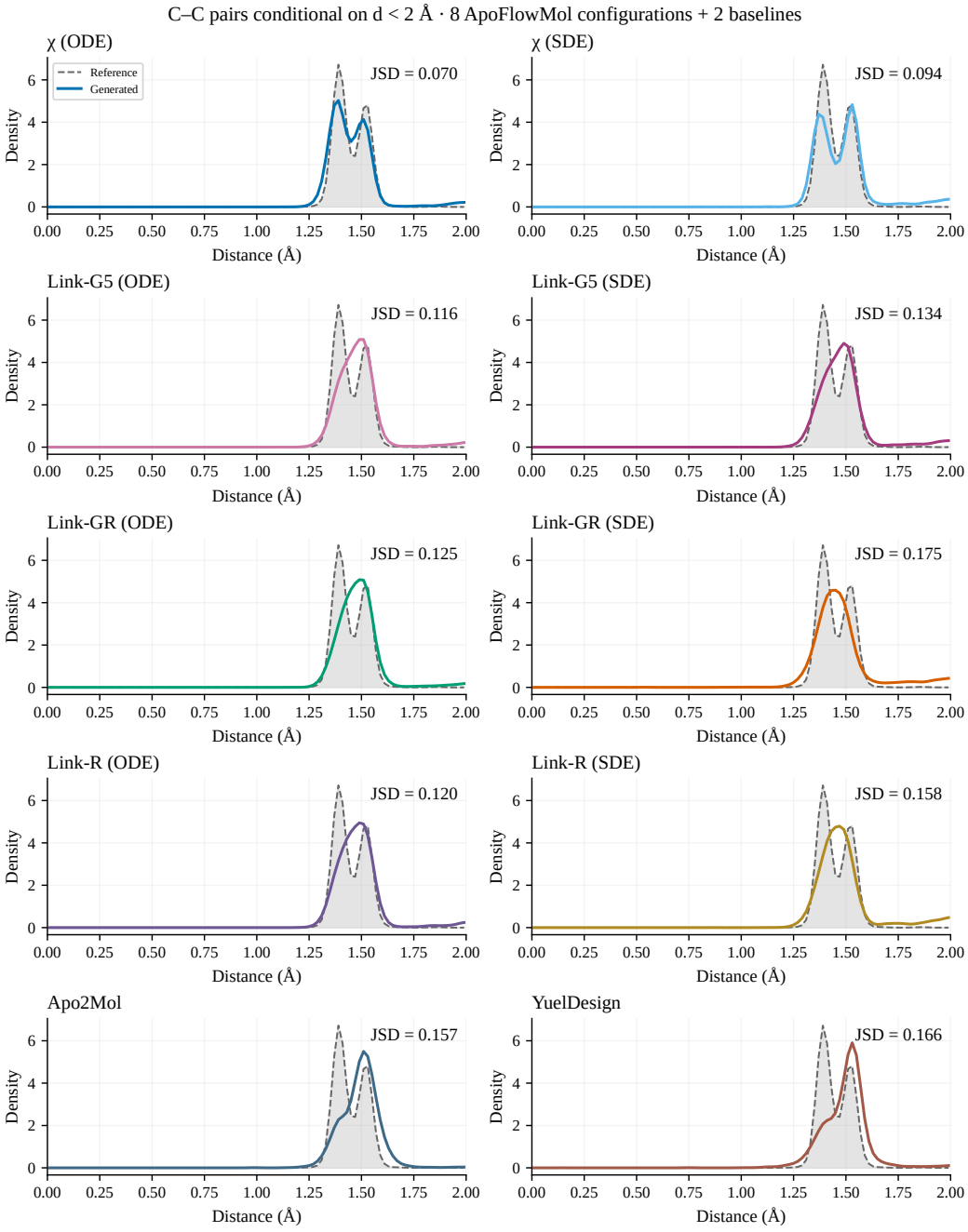


Figure 6: Short-range ligand C–C pair distributions conditioned on $d < 2$ Å.

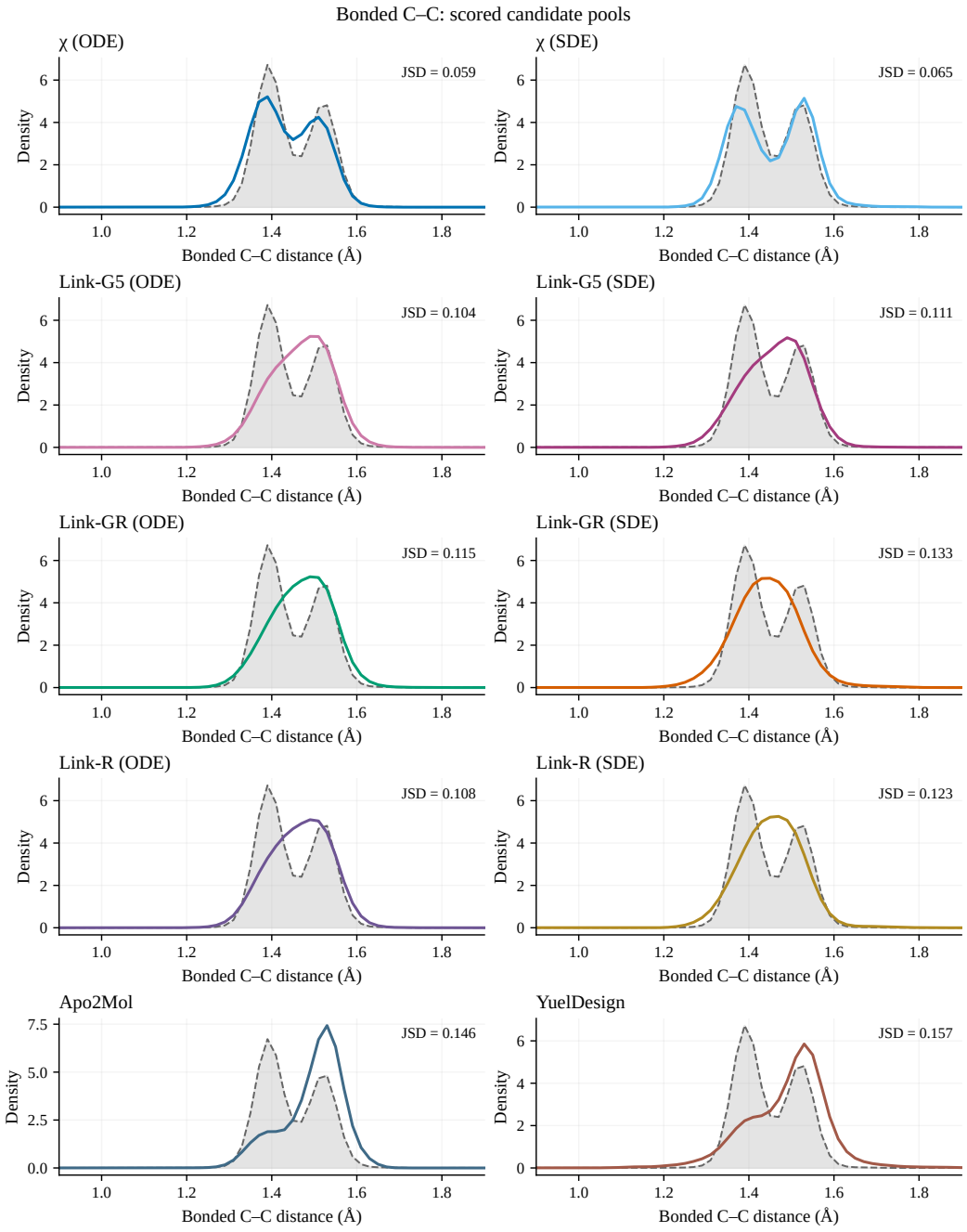


Figure 7: Bonded ligand C–C distributions for the scoring-input pools.

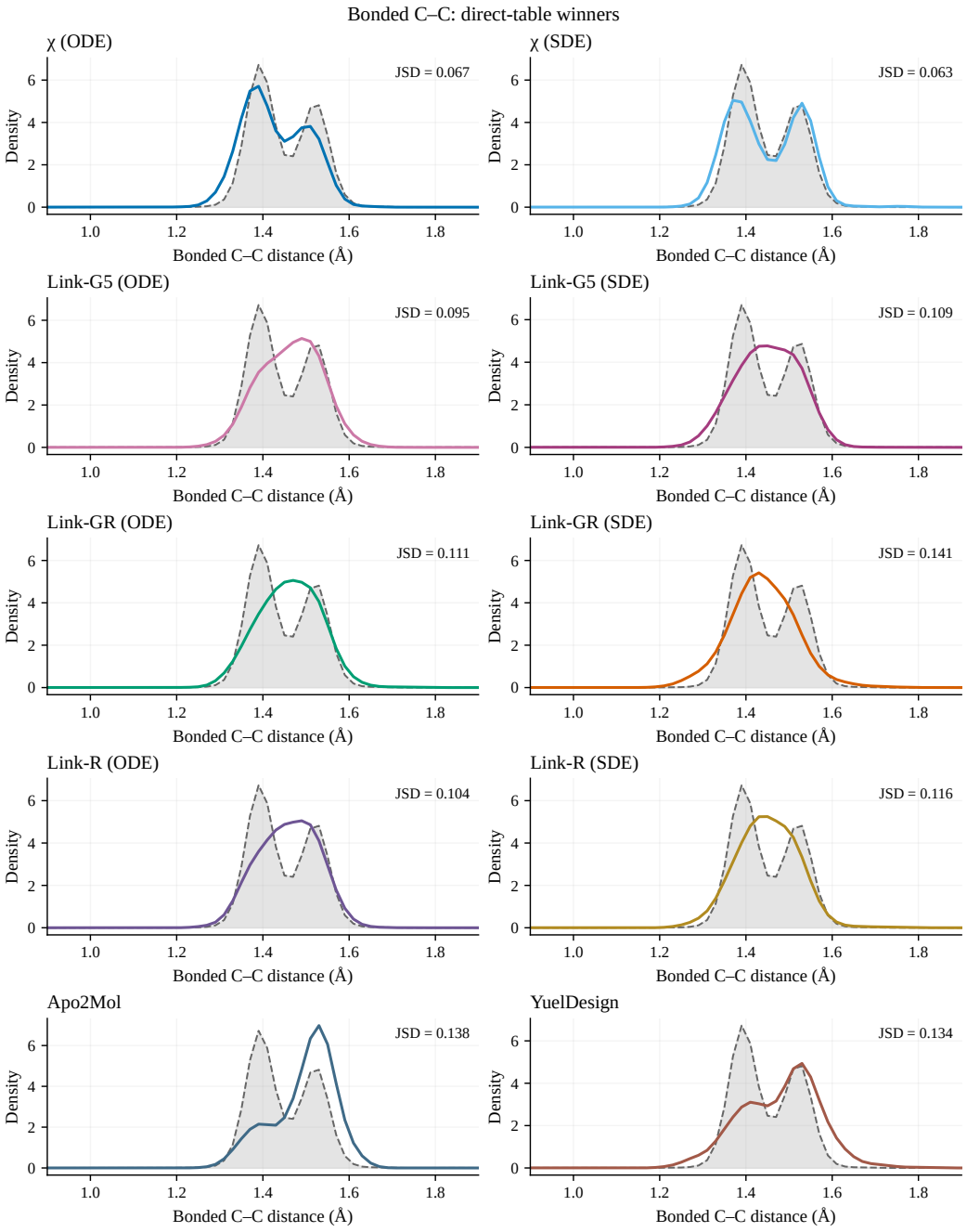


Figure 8: Bonded ligand C–C distributions for the direct-table winners.

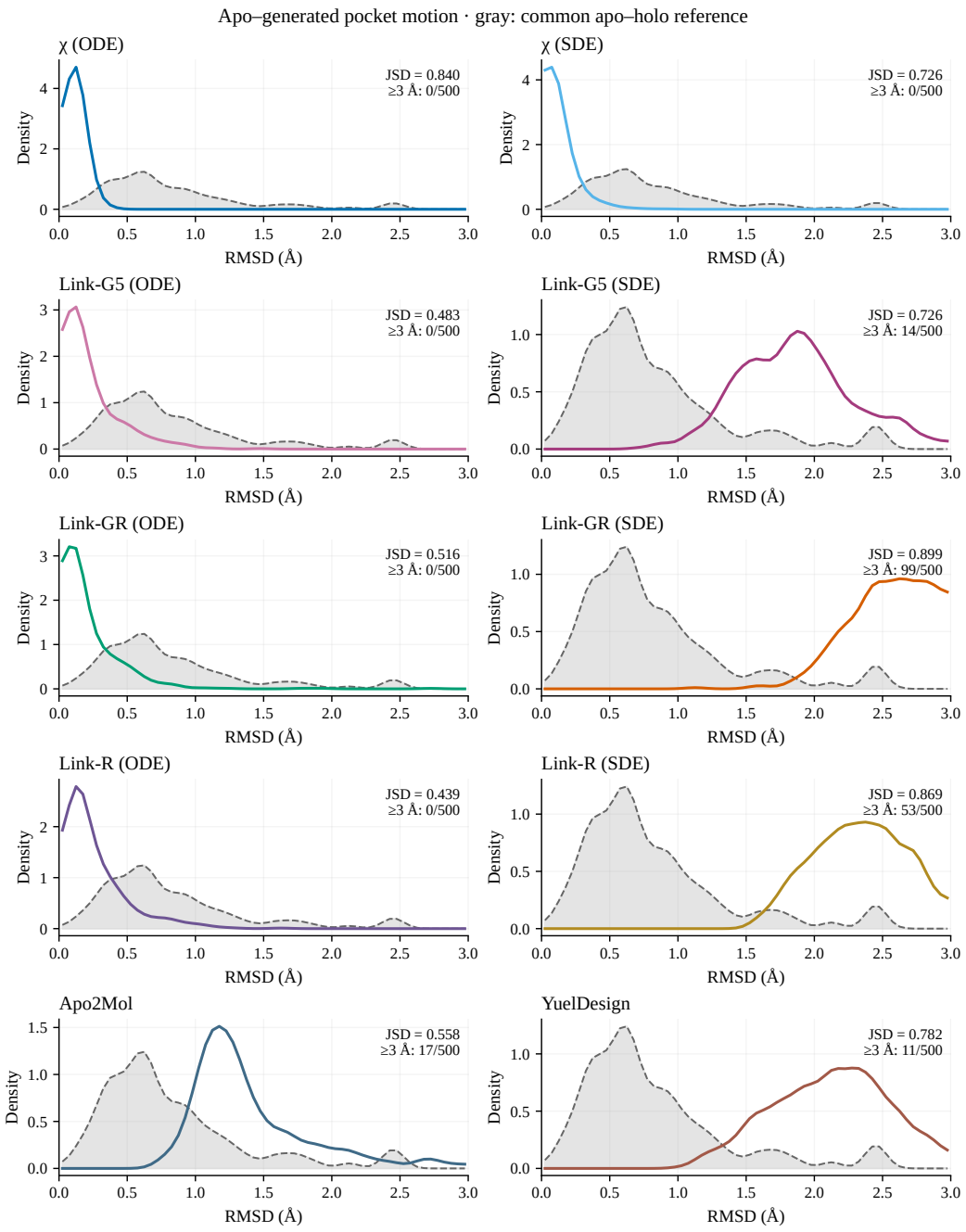


Figure 9: Apo–generated pocket-motion distributions. Gray denotes the matched apo–holo reference; JSD is computed from unsmoothed histograms.

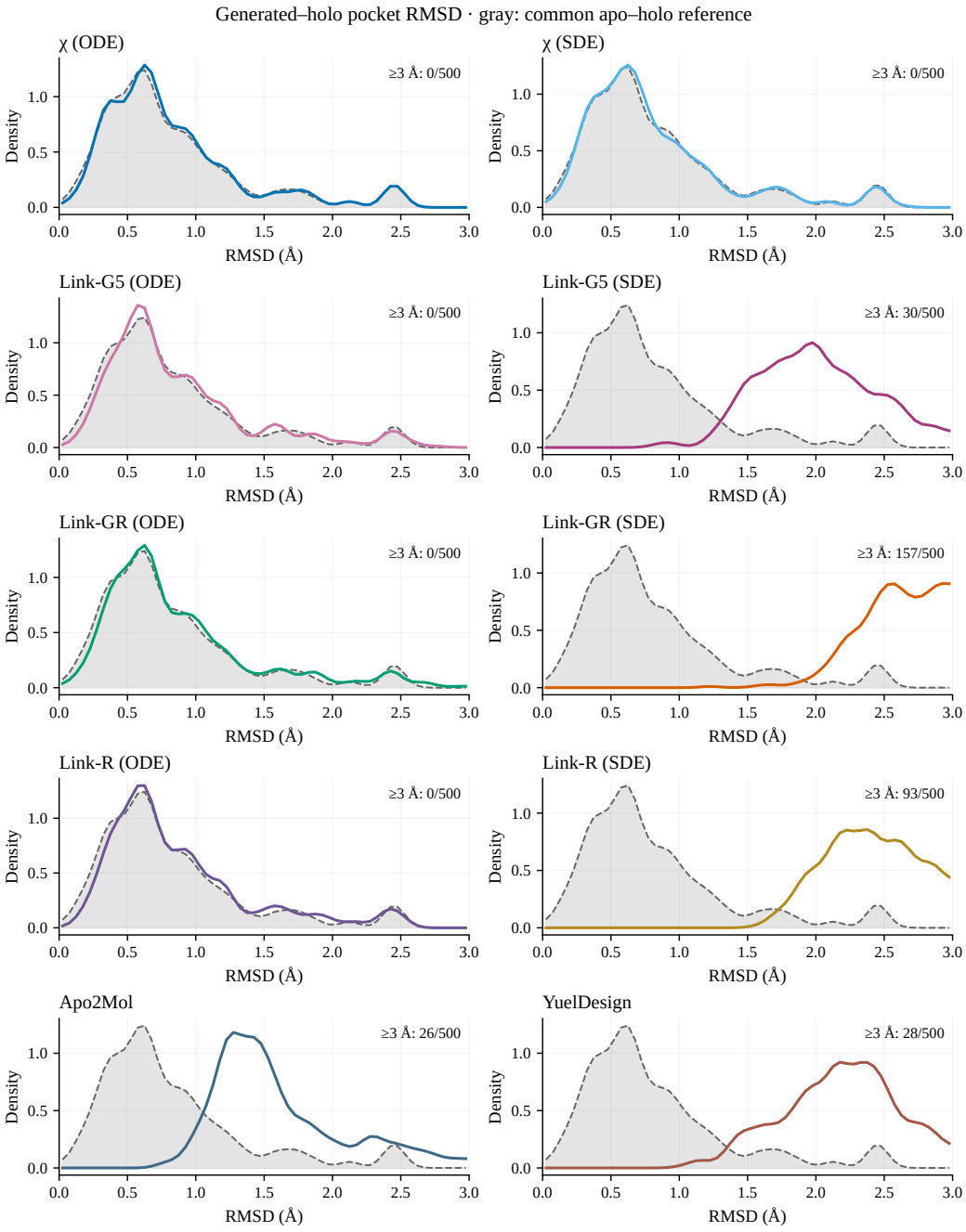


Figure 10: Generated–holo pocket RMSD distributions for all ten methods.

Table 14: Bonded ligand C–C JSD for scoring inputs and direct-table winners.

Method	Scored-pool JSD ↓	Selected-100 JSD ↓
χ (ODE)	0.059	0.067
χ (SDE)	0.065	0.063
Link-G5 (ODE)	0.104	0.095
Link-G5 (SDE)	0.111	0.109
Link-GR (ODE)	0.115	0.111
Link-GR (SDE)	0.133	0.141
Link-R (ODE)	0.108	0.104
Link-R (SDE)	0.123	0.116
Apo2Mol	0.146	0.138
YuelDesign	0.157	0.134

Table 15: Pocket motion on one common heavy-atom correspondence. RMSDs are in angstroms; Avg./Med. use all 500 outputs.

Method	Motion JSD ↓	Apo–gen Avg./Med.	Gen–holo Avg./Med.
χ (ODE)	0.840	0.13/0.12	0.82/0.69
χ (SDE)	0.726	0.13/0.09	0.82/0.68
Link-G5 (ODE)	0.483	0.24/0.15	0.86/0.68
Link-G5 (SDE)	0.726	1.91/1.86	2.08/2.00
Link-GR (ODE)	0.516	0.23/0.14	0.86/0.69
Link-GR (SDE)	0.899	2.72/2.67	2.84/2.80
Link-R (ODE)	0.439	0.27/0.18	0.87/0.69
Link-R (SDE)	0.869	2.43/2.38	2.56/2.50
Apo2Mol	0.558	1.47/1.28	1.69/1.48
YuelDesign	0.782	2.14/2.15	2.23/2.22

G.7 Typed non-covalent contact distributions

We evaluate the native ligand–holo complex and FLOWR-100 conditioned on the original apo pocket, Apo2Mol, YuelDesign, or each of the eight ApoFlowMol configurations, paired over 100 systems. Vina selects ligand identity; contact counts use the original pre-docking generated ligand and its corresponding pocket. We use geometry-based contact definitions. Donor–acceptor candidates use a 3.5 Å distance cutoff; salt-bridge candidates use 4.0 Å; aromatic stacking uses ring-center distance ≤ 5.5 Å and approximately parallel or perpendicular ring normals; cation–π candidates use ionizable feature–ring distances ≤ 6.0 Å. Halogen candidates use a 3.5 Å cutoff and angle at least $140°$. Hydrophobic feature–side-chain C/S pairs use 4.0 Å. Protein backbone C and C_α atoms are excluded from hydrophobic sites. The six-category total sums typed pairs without cross-category deduplication.

Contact counts are discrete. Figures use a common Gaussian kernel of bandwidth one contact only for display; all errors and paired intervals use original integer counts. The ODE and SDE panels separate samplers while showing the same four representations and apo control. Zero marks the native-holo count, not the absence of contacts.

Table 16: Absolute error in total typed-contact counts relative to native holo. Differences and paired 95% intervals compare each condition with apo.

Pocket	MAE	Difference	95% CI	Closer/tied/farther
Apo	5.41	0.00	[0.00,0.00]	0/100/0
Apo2Mol	5.02	-0.39	[-1.36,0.53]	43/9/48
YuelDesign	6.10	0.69	[-0.60,2.00]	41/7/52
χ (ODE)	5.36	-0.05	[-0.90,0.83]	46/19/35
χ (SDE)	4.97	-0.44	[-1.43,0.46]	38/18/44
Link-G5 (ODE)	4.96	-0.45	[-1.39,0.38]	44/14/42
Link-G5 (SDE)	5.03	-0.38	[-1.71,0.92]	49/5/46
Link-GR (ODE)	4.60	-0.81	[-1.73,0.03]	46/15/39
Link-GR (SDE)	5.59	0.18	[-1.12,1.44]	40/7/53
Link-R (ODE)	4.75	-0.66	[-1.65,0.21]	46/9/45
Link-R (SDE)	5.43	0.02	[-1.21,1.29]	44/10/46

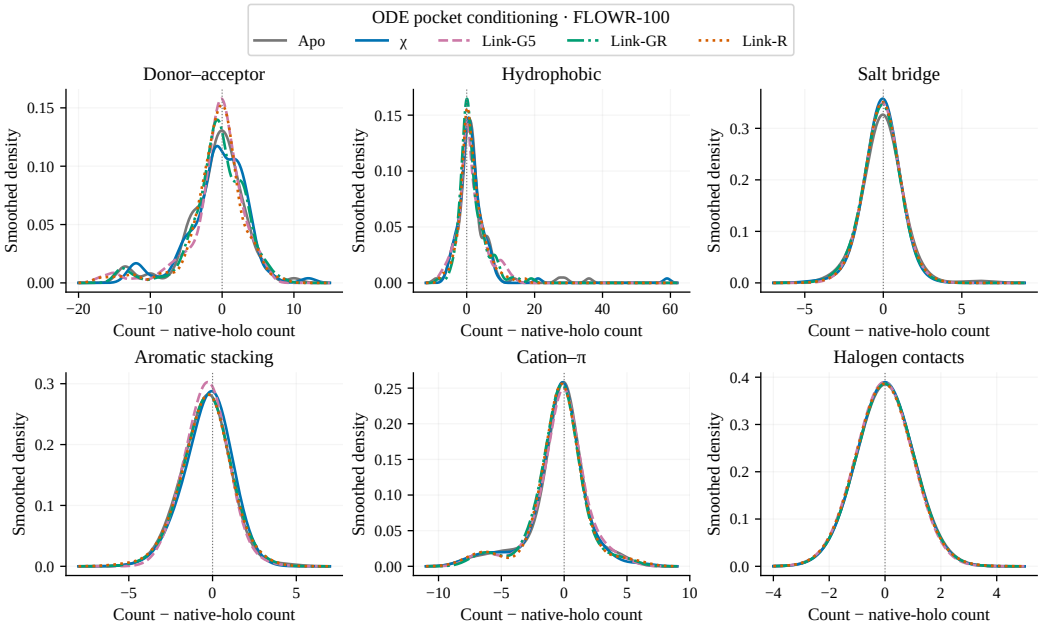


Figure 11: Smoothed per-type contact-count differences for the four ODE configurations and apo control. Each condition contains 100 FLOWR-100 winners.

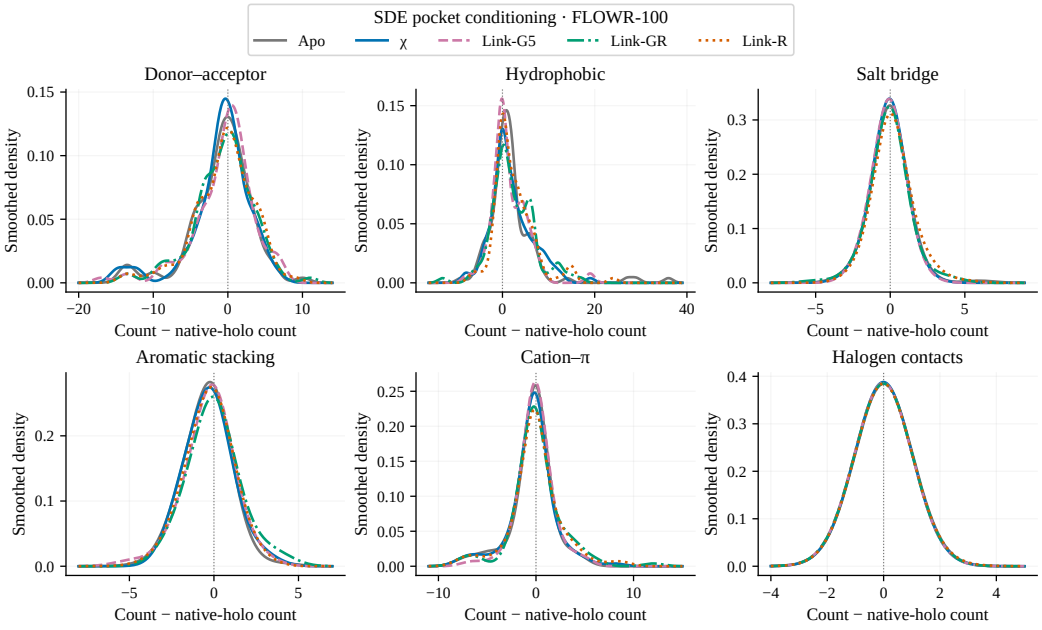


Figure 12: Smoothed per-type contact-count differences for the four SDE configurations and the same apo control. Discrete observations and a shared one-contact bandwidth underlie every curve.

G.8 ILLUSTRATIVE DIRECT-GENERATION EXAMPLES

Figure 4 shows six ApoFlowMol-Link examples together with the matched reference ligand and experimental holo pocket. Table 17 reports the displayed ligand properties and docking scores.

Table 17: Values reported in Figure 4. Vina docking scores are in $\mathrm{kcal\,mol^{-1}}$; SA is normalized so that higher is better.

	QED		SA		Vina ↓	
PDB	Ours	Ref.	Ours	Ref.	Ours	Ref.
7FYI	0.85	0.84	0.80	0.73	-6.70	-5.80
7G0P	0.83	0.37	0.85	0.84	-7.50	-6.80
8PJI	0.78	0.90	0.82	0.86	-6.60	-5.30
7G1M	0.61	0.82	0.55	0.74	-10.30	-9.00
7G1Q	0.66	0.65	0.84	0.77	-9.10	-7.40
8SGE	0.32	0.60	0.58	0.73	-7.80	-6.80

G.9 INTERACTION-LEVEL EXAMPLES

We first examine whether an ApoFlowMol pocket can serve as an effective condition for a separately trained fixed-pocket generator. We freeze the pocket produced by Link-GR-ODE and generate 100 ligands with FLOWR. The two examples in Figure 14, selected from these saved ligands, achieve lower Vina scores and higher QED and normalized SA than their matched reference ligands. For 8PJI, the generated and reference complexes also share the same detected interaction classes: π–π stacking, hydrogen bonding, and hydrophobic interactions.

Figures 15 and 16 provide corresponding direct-generation examples for all eight ApoFlowMol configurations. Each displayed output has a lower Vina docking score than its matched reference ligand.

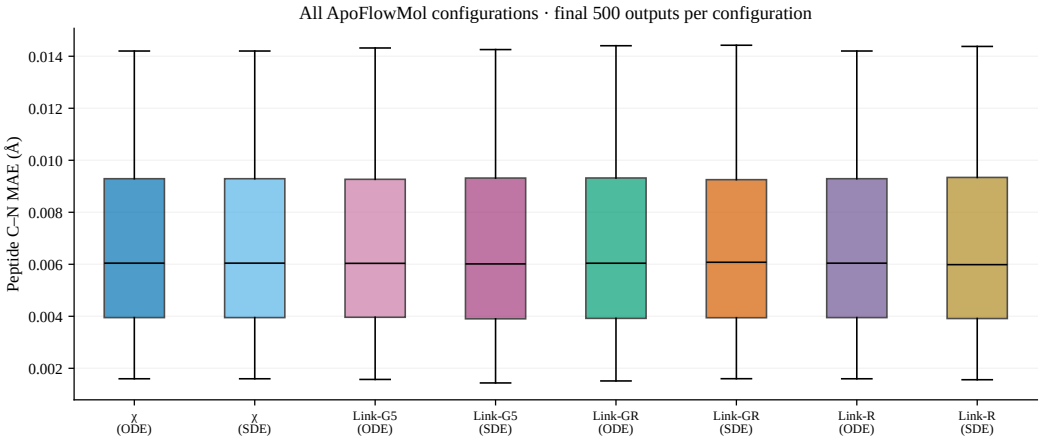
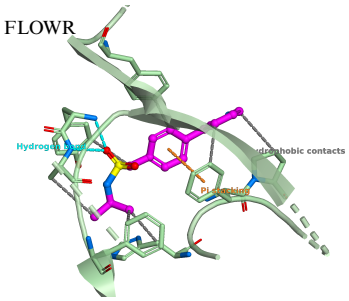
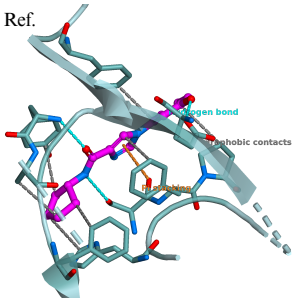


Figure 13: Peptide C–N MAE across all eight ApoFlowMol configurations, measured from the final 500 saved pockets per configuration. Boxes and whiskers use the same convention as the main geometry figure.

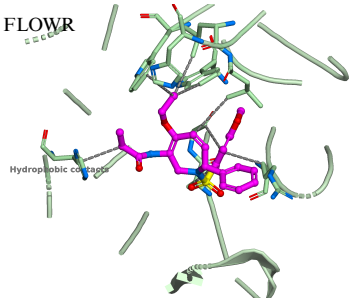
Pocket: Link-GR-ODE		
Ligand: FLOWR		
PDB ID: 8PJI		
	Ours	Ref.
QED	0.92	0.90
SA	0.89	0.86
VINA	-6.30	-5.30
IT	1,2,4	1,2,4





1 π–π stacking 2 Hydrogen bond 3 Salt bridge 4 Hydrophobic interaction 5 Halogen bond

Pocket: Link-GR-ODE		
Ligand: FLOWR		
PDB ID: 8SGE		
	Ours	Ref.
QED	0.67	0.60
SA	0.74	0.73
VINA	-7.70	-6.80
IT	4	2,4



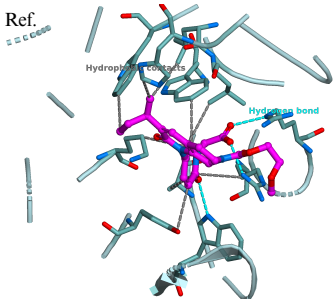
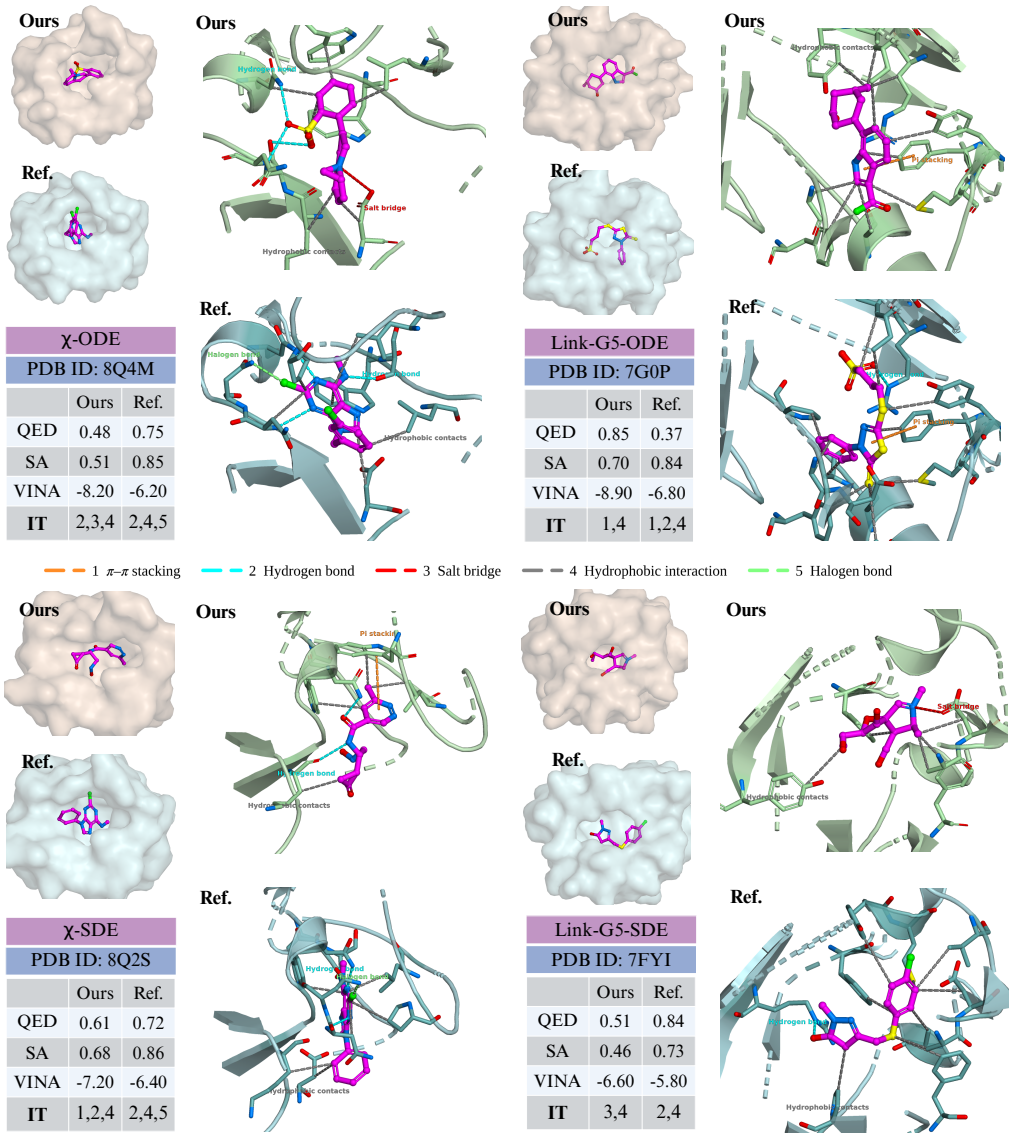


Figure 14: Interaction analysis of FLOWR ligands generated in fixed ApoFlowMol-Link-GR-ODE pockets. "FLOWR" denotes a FLOWR-generated ligand in the fixed ApoFlowMol-adapted pocket; "Ref." denotes the matched reference ligand in its experimental holo pocket. QED and normalized SA are reported together with Vina docking scores in $\mathrm{kcal\,mol^{-1}}$. IT denotes the detected interaction types: 1, π–π stacking; 2, hydrogen bond; 3, salt bridge; 4, hydrophobic interaction; and 5, halogen bond.



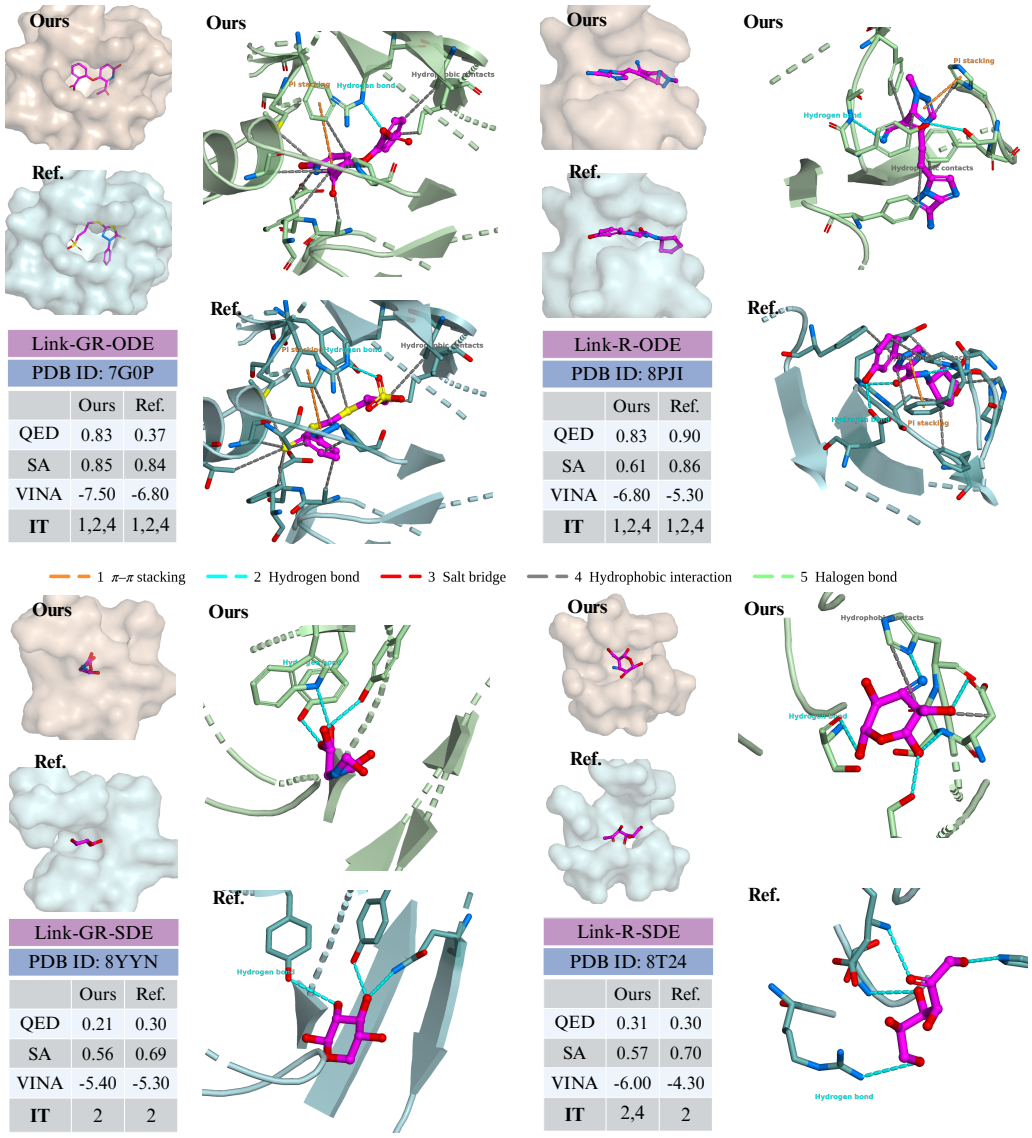
χ-ODE		
PDB ID: 8Q4M		
	Ours	Ref.
QED	0.48	0.75
SA	0.51	0.85
VINA	-8.20	-6.20
IT	2,3,4	2,4,5

Link-G5-ODE		
PDB ID: 7G0P		
	Ours	Ref.
QED	0.85	0.37
SA	0.70	0.84
VINA	-8.90	-6.80
IT	1,4	1,2,4

χ-SDE		
PDB ID: 8Q2S		
	Ours	Ref.
QED	0.61	0.72
SA	0.68	0.86
VINA	-7.20	-6.40
IT	1,2,4	2,4,5

Link-G5-SDE		
PDB ID: 7FYI		
	Ours	Ref.
QED	0.51	0.84
SA	0.46	0.73
VINA	-6.60	-5.80
IT	3,4	2,4

Figure 15: Selected direct-generation examples for configurations 1–4: χ-ODE, χ-SDE, Link-G5-ODE, and Link-G5-SDE. "Ours" denotes the generated ligand–pocket pair, and "Ref." denotes the matched ground-truth ligand and experimental holo pocket. QED, normalized SA, and Vina docking scores are shown; IT follows the interaction-type numbering in Figure 14.



Link-GR-ODE		
PDB ID: 7G0P		
	Ours	Ref.
QED	0.83	0.37
SA	0.85	0.84
VINA	-7.50	-6.80
IT	1,2,4	1,2,4

Link-R-ODE		
PDB ID: 8PJI		
	Ours	Ref.
QED	0.83	0.90
SA	0.61	0.86
VINA	-6.80	-5.30
IT	1,2,4	1,2,4

Link-GR-SDE		
PDB ID: 8YYN		
	Ours	Ref.
QED	0.21	0.30
SA	0.56	0.69
VINA	-5.40	-5.30
IT	2	2

Link-R-SDE		
PDB ID: 8T24		
	Ours	Ref.
QED	0.31	0.30
SA	0.57	0.70
VINA	-6.00	-4.30
IT	2,4	2

Figure 16: Selected direct-generation examples for configurations 5–8: Link-GR-ODE, Link-GR-SDE, Link-R-ODE, and Link-R-SDE. Labels, metrics, and interaction-type numbering follow Figure 14.