TSEA: A LARGE-SCALE DATASET OF SYNTHETIC TERNARY COMPLEXES FOR MOLECULAR GLUE MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Molecular glues (MGs) are an emerging therapeutic modality that induces or stabilizes protein–protein interactions, with demonstrated clinical impact in oncology. However, the scarcity of experimentally resolved MG ternary structures severely limits data-driven molecular glue modeling. Here, we introduce **TSea**, a large-scale dataset comprising over 100,000 synthetic ternary-complex structures, more than 450 times the size of MG-PDB. Starting from experimentally determined protein–ligand complexes, we construct synthetic ternary complexes through structurally constrained protein-chain partitioning and contact-based screening against MG-PDB reference distributions. Comparative analyses show that TSea captures selected structural and interaction characteristics of MG-PDB. We further develop **TSeaDiff**, a diffusion-based molecular generator trained on TSea. Among the evaluated models, TSeaDiff achieves the closest agreement with MG-PDB in protein–ligand interaction profiles. These results support TSea as a scalable structural resource for data-driven molecular glue modeling.

1 INTRODUCTION

Molecular glues (MGs) have emerged as an important therapeutic modality for cancer treatment, with drugs already approved for clinical use (King et al., 2026). Whereas many conventional small-molecule drugs modulate an individual protein, molecular glues induce or stabilize protein–protein interactions (PPIs). These MG-induced interactions can trigger biological events such as targeted protein degradation or altered protein activity, providing therapeutic mechanisms beyond conventional inhibition (Mayor-Ruiz et al., 2020; Holderfield et al., 2024).

Despite their therapeutic promise, molecular glue discovery has historically relied heavily on serendipitous findings. More systematic discovery strategies, including scalable chemical profiling, have begun to provide more directed routes to identifying MG drugs (Mayor-Ruiz et al., 2020). However, the scarcity of experimentally resolved MG-mediated ternary complexes severely limits the application of structure-based machine learning to MGs (Xue et al., 2025). Although binary protein–ligand structures provide abundant binding information, they do not capture ligand binding at an interface jointly formed by two protein partners. MG-PDB, which curates experimentally resolved non-covalent MG ternary complexes, contains only 221 structures, leaving limited chemical and structural coverage for data-driven modeling. The difficulty is also evident in ternary-structure prediction: on MGBench, AlphaFold 3 (AF3) achieved a success rate of only 32.9% in recovering molecular glue–protein interactions (Liao et al., 2025).

Existing efforts have expanded ternary structural resources through database mining. For example, TernaryDB extracts protein–ligand–protein complexes from ligand-containing multichain protein structures in the PDB (Xue et al., 2025). Although this approach increases structural coverage, its scale and diversity remain constrained by the availability of such experimentally resolved structures. Moreover, distance-based mining criteria alone do not ensure that retrieved structures exhibit the ligand-contact patterns and interfacial geometries observed in experimental MG complexes. These limitations motivate constructing synthetic ternary complexes from a broader pool of binary protein–ligand structures, guided by experimental MG binding characteristics.

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 these limitations, we introduce **TSea**, a large-scale dataset of synthetic ternary complexes constructed by partitioning experimentally resolved binary protein–ligand structures. We systematically analyze the ligand properties and intermolecular interaction profiles of experimentally resolved non-covalent MG-mediated ternary systems and use these observations to guide binary-complex selection and protein partitioning. Guided by these criteria, selected protein chains are partitioned into two components that jointly engage the bound ligand while satisfying structural and contact constraints. Applying this construction pipeline at scale yields over 100,000 synthetic ternary structures, substantially expanding the structural data available for modeling ligand-mediated protein–protein interfaces. We further assess their agreement with experimental MG complexes using complementary structural, scoring, and intermolecular interaction descriptors.

To investigate the use of TSea for molecular generation, we train **TSeaDiff**, a diffusion model for ligand generation conditioned on two fixed protein pockets, using pairwise protein–protein and ligand–protein interaction modeling to guide ligand denoising. We compare TSeaDiff with existing molecular generators on held-out, locally perturbed TSea pockets. The evaluation examines whether the generated ligands reproduce structural, scoring, and interaction characteristics observed in experimental MG complexes.

To summarize, our main contributions are:

- We systematically characterize experimentally resolved non-covalent molecular glue complexes in MG-PDB, quantifying ligand properties and interfacial contact patterns to establish an empirical structural reference for synthetic ternary-complex curation.
- We introduce **TSea**, a dataset of over 100,000 synthetic ternary complexes, together with a reusable and scalable curation pipeline applicable to different sources of binary protein–ligand structures. Comparative analyses show that TSea captures selected structural and interaction characteristics of experimental molecular glue complexes.
- We develop **TSeaDiff**, a diffusion model for ligand generation conditioned on protein–protein interface pockets, using pairwise protein–protein and ligand–protein interaction modeling to guide ligand denoising. We compare TSeaDiff with existing molecular generators on locally perturbed, held-out test pockets and evaluate whether the generated ligands reproduce structural, scoring, and interaction characteristics observed in experimental molecular glue complexes.

2 Related Work

Molecular-glue ternary structural resources. MG-PDB curates experimentally resolved non-covalent MG ternary complexes, and MGBench provides a benchmark subset for ternary structure prediction (Liao et al., 2025). MGTbind further integrates MG interaction records, experimental structures, predicted ternary complexes, and covalently bound MG systems (Zhu et al., 2026). TSea uses MG-PDB-derived ligand and contact statistics to guide synthetic ternary-complex construction.

Large-scale ternary structural data. TernaryDB expands available training data by mining protein–ligand–protein complexes from the PDB using structural quality and geometric contact criteria (Xue et al., 2025). TSea complements this approach by partitioning binary protein–ligand structures into synthetic ternary environments under structural constraints and MG-derived contact criteria.

Pocket-conditioned molecular generation. Recent structure-based molecular generation models, including TargetDiff (Guan et al., 2023), DiffSBDD (Schneuing et al., 2024), PocketXMol (Peng et al., 2026), AnewOmni (Kong et al., 2026), and ODesign (Zhang et al., 2025), generate small molecules conditioned on three-dimensional protein environments. For ligand generation at ternary interfaces, our TSeaDiff model combines pairwise protein–protein and ligand–protein interaction modeling with joint ternary denoising.

3 Structural Analysis of MG-PDB

Given the limited availability of experimentally resolved molecular glue (MG)-bound ternary complexes, characterizing existing structures is essential for guiding the construction of TSea. We

therefore analyzed intermolecular atomic contacts and ligand properties in MG-PDB to establish an experimental structural reference for dataset construction.

3.1 Intermolecular Atomic Contact Analysis

MG-mediated ternary complexes with an interfacial binding mode involve protein–protein contacts and contacts between the MG ligand and each protein partner. We characterized these three interfaces to quantify the contact patterns of the ternary binding environment.

Let A and B denote the two protein components and L the ligand. Their corresponding heavy-atom sets are denoted by $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{L}$, respectively. We use AB, AL, and BL as labels for the corresponding pairwise interfaces. Two atoms from different components are considered to be in contact if their Euclidean distance is at most $d_c = 4.0\,\text{Å}$, a cutoff commonly used to identify close intermolecular contacts in structural analyses (Khazanov & Carlson, 2013; Kortemme & Baker, 2002).

For two disjoint heavy-atom sets $\mathbb{X}$ and $\mathbb{Y}$, we define the directed contact-atom count as

$$N_{\mathbb{X}\to\mathbb{Y}} = \left|\{y \in \mathbb{Y} \,|\, \exists\, x \in \mathbb{X} : \|\mathbf{r}_x - \mathbf{r}_y\|_2 \le d_c\}\right|, \tag{1}$$

where $\mathbf{r}_x, \mathbf{r}_y \in \mathbb{R}^3$ are atomic coordinates. This quantity counts the distinct heavy atoms in $\mathbb{Y}$ that contact at least one heavy atom in $\mathbb{X}$. Each target atom is counted only once, and the arrow specifies the counting direction rather than a physical directionality of the interaction. Applying this definition in both directions at each of the three interfaces characterizes the contacting atoms on both sides of each interface.

To additionally capture the multiplicity of atomic contacts, we define the contact-pair count as

$$C_{\mathbb{X},\mathbb{Y}} = \left|\{(x,y) \in \mathbb{X}\times\mathbb{Y} \,|\, \|\mathbf{r}_x - \mathbf{r}_y\|_2 \le d_c\}\right|. \tag{2}$$

Each contacting atom pair is counted once, while an individual atom may participate in multiple pairs. This count is symmetric: $C_{\mathbb{X},\mathbb{Y}} = C_{\mathbb{Y},\mathbb{X}}$.

We summarized each ternary complex using two complementary contact descriptors:

$$\begin{aligned} \overline{N}_{\mathrm{PP}} &= \frac{N_{\mathbb{A}\to\mathbb{B}} + N_{\mathbb{B}\to\mathbb{A}}}{2}, \\ \overline{C}_{\mathrm{LP}} &= \frac{C_{\mathbb{L},\mathbb{A}} + C_{\mathbb{L},\mathbb{B}}}{2}. \end{aligned} \tag{3}$$

Here, $\overline{N}_{\mathrm{PP}}$ measures the mean number of distinct contacting heavy atoms on the two sides of the protein–protein interface, whereas $\overline{C}_{\mathrm{LP}}$ measures the mean number of heavy-atom contact pairs across the two ligand–protein interfaces. The former characterizes the extent of the protein–protein interface in terms of participating atoms, while the latter captures the multiplicity of ligand–protein contacts.

For each of the $n = 221$ analyzed MG-PDB structures, the two contact descriptors $\overline{C}_{\mathrm{LP}}$ and $\overline{N}_{\mathrm{PP}}$ were standardized using their respective reference-set means and standard deviations. The resulting standardized descriptor vector for structure i is denoted by $\mathbf{z}_i$. To establish a joint reference distribution for TSea construction, we fitted a bivariate Gaussian kernel density estimate (KDE) (Silverman, 1986):

$$\widehat{p}_h(\mathbf{z}) = \frac{1}{2\pi n h^2} \sum_{i=1}^{n} \exp\left(-\frac{\|\mathbf{z} - \mathbf{z}_i\|_2^2}{2h^2}\right), \qquad \mathbf{z} \in \mathbb{R}^2. \tag{4}$$

Here, $h > 0$ is the bandwidth of the isotropic Gaussian kernel in the standardized descriptor space, and each reference structure contributes equal weight $1/n$. The bandwidth was selected by leave-one-out (LOO) likelihood cross-validation. The resulting KDE provides a smooth joint reference distribution of $\overline{C}_{\mathrm{LP}}$ and $\overline{N}_{\mathrm{PP}}$ without assuming a specific parametric form for their joint distribution.

From the fitted reference density, we derived the 95% and 99% highest density regions (HDRs), defined as density superlevel sets containing the corresponding probability mass under the KDE (Hyndman, 1996). These nested regions (Fig. 1) provide stricter and more inclusive contact-based screening criteria, respectively, for identifying candidates with contact patterns supported by the experimental reference distribution. Feature standardization, bandwidth selection, and HDR-based screening are detailed in Appendix A.

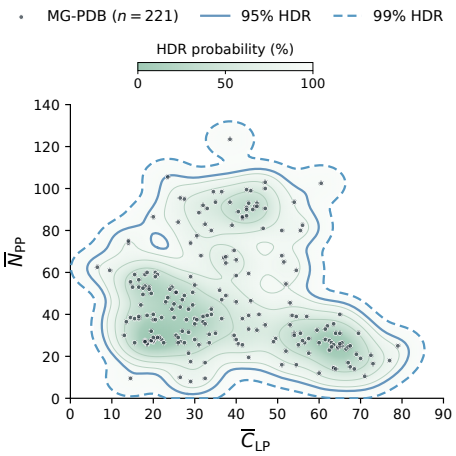


Figure 1: **MG-PDB contact distribution and highest density regions**

3.2 MG Ligand Analysis

We further characterized MG ligands by their size, protein-contact coverage, and balance of contacts with the two protein partners. These ligand-level properties complement the joint contact analysis and inform the screening criteria used for TSea construction. Detailed results are provided in Appendix B.

Ligand size. MG ligands in the analyzed MG-PDB structures contained 10–70 heavy atoms. To accommodate modest variation beyond the observed range, we used the broader range $10 \leq |\mathbb{L}| < 75$ heavy atoms for TSea curation.

Ligand contact coverage. To quantify the fraction of ligand atoms contacting either protein component, we define ligand contact coverage as

$$f_{\mathrm{contact}} = \frac{N_{\mathbb{P}\to\mathbb{L}}}{|\mathbb{L}|}, \qquad \mathbb{P} = \mathbb{A} \cup \mathbb{B}. \tag{5}$$

Each ligand heavy atom is counted only once, including atoms contacting both components.

Ligand–protein contact centrality. To complement overall contact coverage with a measure of contact balance between the two protein components, we define ligand–protein contact centrality as

$$c_{\mathrm{LP}} = 1 - \frac{|N_{\mathbb{A}\to\mathbb{L}} - N_{\mathbb{B}\to\mathbb{L}}|}{N_{\mathbb{A}\to\mathbb{L}} + N_{\mathbb{B}\to\mathbb{L}}}. \tag{6}$$

Each component-specific count includes the distinct ligand heavy atoms contacting that component; an atom contacting both proteins contributes once to each count. For complexes with at least one ligand–protein contact, $c_{\mathrm{LP}} \in [0, 1]$, with 1 indicating equal contact-atom counts for the two components and 0 indicating contact with only one component. Both f_{contact} and c_{LP} are invariant to exchanging $\mathbb{A}$ and $\mathbb{B}$.

4 Dataset Construction and Evaluation

The preceding analyses provided the empirical basis for the ligand and contact criteria used to construct TSea. We applied these criteria, together with constraints on protein secondary structure and fragment length, to identify suitable partitions of binary protein–ligand complexes into synthetic ternary structures (Fig. 2).

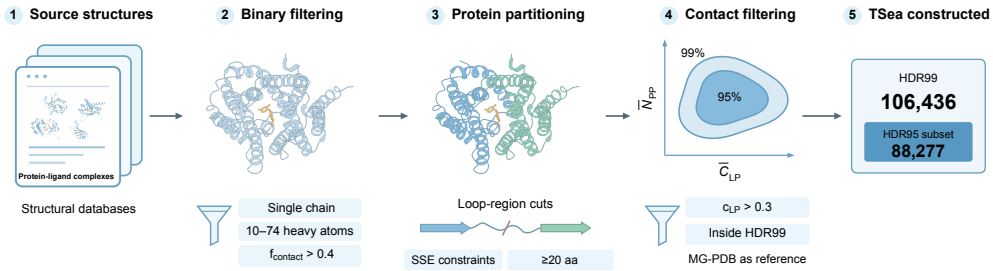


Figure 2: **TSea construction pipeline.** Binary protein–ligand complexes undergo filtering, protein partitioning, and MG-PDB–referenced contact selection.

4.1 Binary Complex Filtering and Protein Partitioning

Source-complex filtering. We collected candidate complexes from the regular small-molecule subset of BioLiP2, a curated database of biologically relevant protein–ligand interactions derived from experimental PDB structures (Zhang et al., 2024). To reduce potential artifacts, we further applied a blacklist of Chemical Component Dictionary (CCD) identifiers targeting common crystallization additives, purification-buffer components, and other potentially non-biological ligands (Yang et al., 2013). Eligible source complexes contained a single protein chain and a ligand with 10–74 heavy atoms, with ligand contact coverage $f_{\text{contact}} > 0.4$. We also excluded complexes with short ligand–protein contacts suggestive of covalent linkage or steric clashes.

Protein partitioning and contact-based selection. For each retained source complex, we evaluated candidate partition sites in loop regions and applied constraints on cross-fragment β-sheet connectivity to limit disruption of secondary-structure organization. Each resulting protein fragment was required to contain at least 20 residues. Candidate partitions were further required to satisfy $c_{\text{LP}} > 0.3$ and to fall within the 99% highest density region (HDR99) of the MG-PDB reference distribution in the $(\overline{C}_{\text{LP}}, \overline{N}_{\text{PP}})$ contact space defined in Section 3.

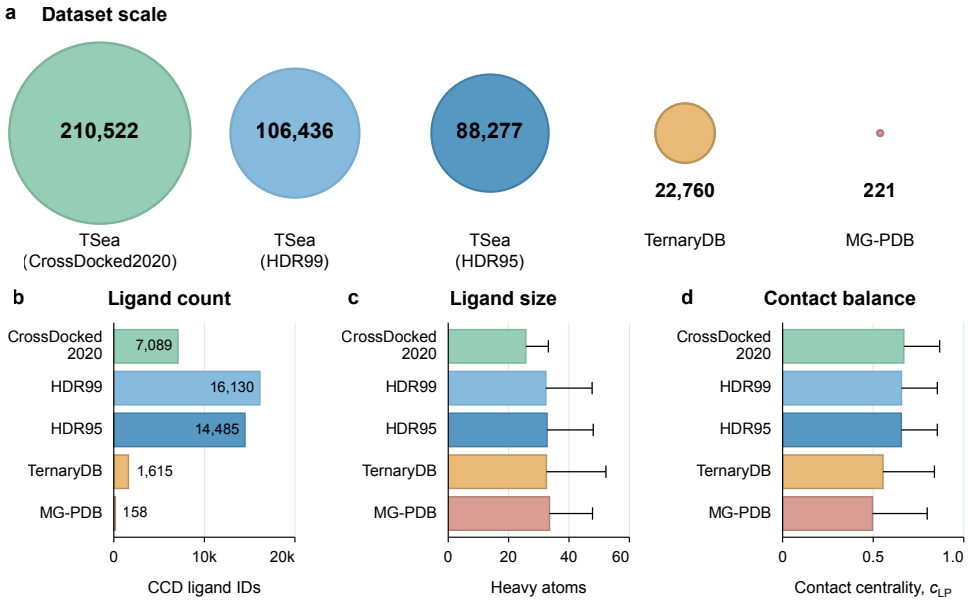


Figure 3: **Dataset scale and ligand properties.** **(a)** Complex counts, represented by circle area. **(b)** CCD ligand counts. **(c)** Ligand heavy-atom counts. **(d)** Contact centrality, c_{LP}.

4.2 TSEA DATASET EVALUATION

Following dataset construction, we evaluated TSea from complementary perspectives of dataset scale and ligand coverage, fixed-pose AutoDock Vina scoring (Trott & Olson, 2010), non-covalent interactions identified by the Protein–Ligand Interaction Profiler (PLIP, Schake et al., 2025), and buried surface area (BSA), using TernaryDB and MG-PDB as reference datasets. Complete metric definitions, statistics, and additional analyses are provided in Appendix D.

Scale and ligand coverage. TSea contains 106,436 structures in HDR99 and 88,277 in the nested HDR95 subset, corresponding to approximately 482-fold and 399-fold the number of experimentally resolved structures in MG-PDB, respectively. The two variants span 16,130 and 14,485 CCD ligand identifiers, substantially expanding both structural and ligand coverage beyond existing ternary-complex datasets while maintaining ligand-size distributions close to the experimental MG reference (Fig. 3).

Vina scoring. We characterized ligand–protein scoring profiles using fixed-pose AutoDock Vina scores for the combined protein partners and for each partner separately, together with Vina-score centrality (Fig. 4). TSea at both HDR levels exhibited L–AB and stronger-side scores that were closer to MG-PDB than those of TernaryDB, although they remained less negative than the experimental reference. The weaker-side scores showed particularly close agreement with MG-PDB and were substantially more negative than those of TernaryDB, which is particularly important for ternary-complex formation (Chen & Zacharias, 2023). TSea also exhibited higher score centrality than both MG-PDB and TernaryDB, indicating more balanced score magnitudes across the two protein partners among structures with favorable scores on both sides.

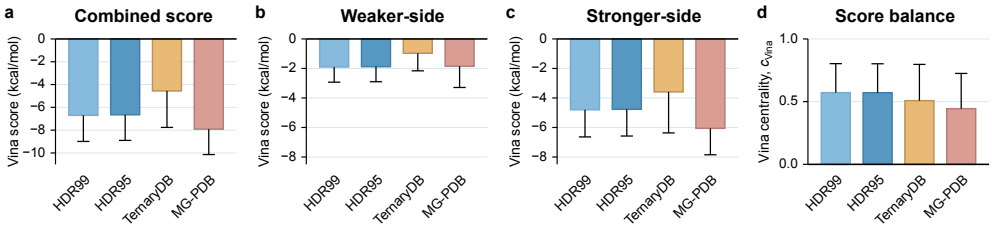


Figure 4: **Fixed-pose Vina scoring.** **(a)** Combined, **(b)** weaker-side, and **(c)** stronger-side scores. **(d)** Vina centrality, c_{Vina}.

Intermolecular interactions. We further used PLIP to characterize non-covalent protein–protein and protein–ligand interactions. Agreement with MG-PDB was quantified using a reference-standardized root-mean-square deviation over five shared interaction types, with lower values indicating greater similarity (Fig. 5). TSea showed smaller PPI and PLI profile distances to MG-PDB than TernaryDB, while its protein–ligand interaction density and interaction-type richness were also closer to the experimental reference. At the synthetic protein–protein interface, hydrophobic interactions were more prevalent, as expected from bringing part of the hydrophobic protein interior to the newly defined interfacial surfaces through chain partitioning.

Buried surface area. BSA analysis provided a complementary geometric assessment of the constructed interfaces (Fig. 6). At ligand-centered pocket scales, TSea reproduced the mean local total buried surface area of MG-PDB more closely than TernaryDB, and the mean full-structure protein–ligand interface BSAs, particularly that of the smaller of the two ligand-contacting sides, were close to the experimental reference. TSea also exhibited substantially narrower total BSA distributions than TernaryDB, with both the mean values and dispersion more closely matching MG-PDB at ligand-centered pocket scales.

Together, the analyses above demonstrate that the proposed curation pipeline can construct large-scale synthetic ternary complexes whose structural and interaction distributions resemble those of experimentally resolved MG-mediated ternary complexes. As an additional construction ablation, removing HDR screening increased the deviation from MG-PDB in descriptors not used to define

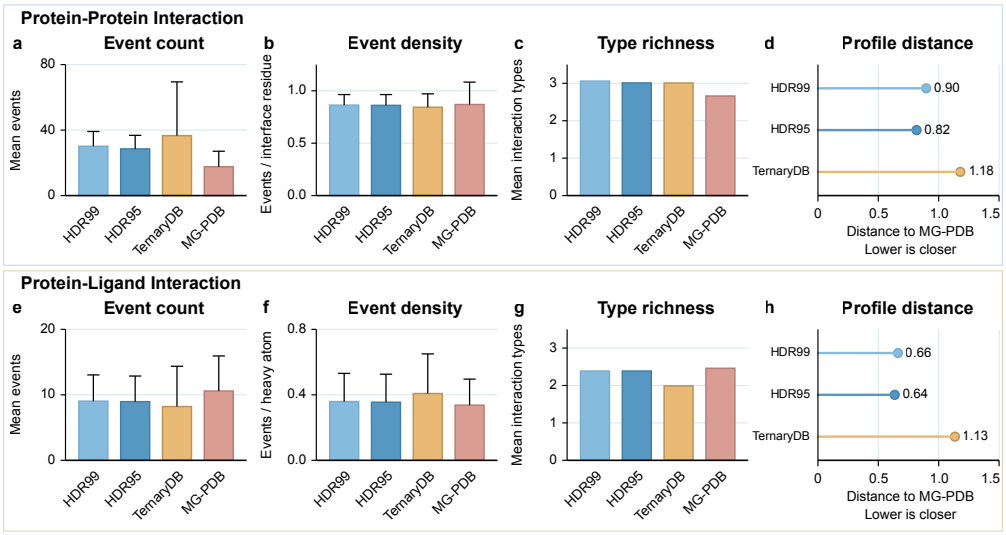


Figure 5: **Full-dataset PLIP interaction profiles.** PPI (top) and PLI (bottom): **(a, e)** event counts; **(b, f)** interaction densities; **(c, g)** interaction-type richness; **(d, h)** MG-PDB–standardized profile distances. Lower distances indicate closer agreement.

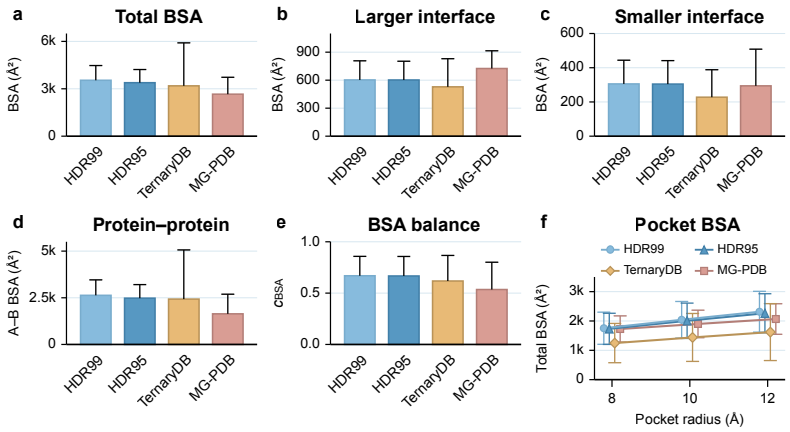


Figure 6: **Buried surface area (BSA).** **(a)** Total BSA. **(b)** Larger and **(c)** smaller protein–ligand interfaces. **(d)** Protein–protein interface. **(e)** BSA centrality, c_{BSA}. **(f)** Pocket BSA; markers are horizontally offset for clarity.

the HDR criterion, including PLIP interaction profiles and protein–protein interfacial BSA (Appendix D.5), supporting the role of HDR screening in constraining the resulting distribution beyond the descriptors directly used for selection. Moreover, the pipeline is not restricted to BioLiP2 and can also be applied to other sources of binary protein–ligand complexes. For instance, applying the same strategy to the CrossDocked2020 dataset (Francoeur et al., 2020) yielded an additional large-scale ternary complex collection (Fig. 3a), demonstrating the broader applicability of the TSea curation pipeline.

5 Downstream Application

5.1 Molecular generation model

TSea provides training examples in which a small molecule engages both sides of a protein–protein interface. We use these examples to train **TSeaDiff**, a conditional molecular generator for PPI

interface pockets (Fig. 7). Our aim is to learn the geometric and chemical features of ligand interactions with both protein partners, and to assess whether a generator trained on TSea can reproduce molecular-glue-like interaction patterns.

To capture these interactions, we introduce a *pairwise-to-ternary* denoising architecture inspired by alternative routes to MG-mediated ternary complex formation: protein–protein association followed by glue binding, or ligand binding to one protein followed by recruitment of the other (Rui et al., 2023). At each denoising step, a protein–protein expert encodes the A–B context, while a parameter-shared binding expert separately processes the ligand with A and with B. These pairwise representations are combined through symmetric residual averaging and jointly refined by a common equivariant block. The resulting design integrates protein–protein and ligand–protein context while treating the two protein partners symmetrically. It encodes an interaction-based inductive bias without prescribing a dominant binding route or modeling binding kinetics. Architecture and training details are provided in Appendix E.

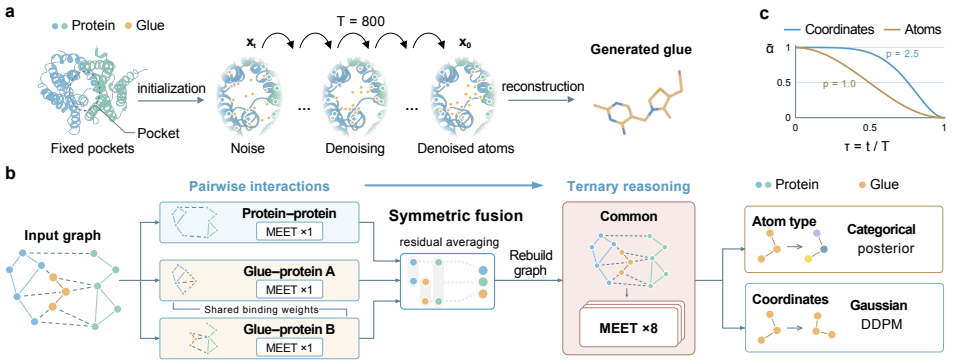


Figure 7: **TSeaDiff architecture. (a)** Molecular glue generation conditioned on fixed protein pockets. **(b)** Pairwise interaction experts, symmetric fusion, and ternary graph refinement for atom-type and coordinate denoising. **(c)** Noise schedules for coordinates and atom types.

5.2 Generation Results

We compared TSeaDiff with existing structure-based molecular generators on 100 locally perturbed pockets from the held-out test partition of a scaffold- and source-group-based split of TSea. The pockets were constructed through energy-constrained side-chain repacking and selected using a ligand-proximity-weighted displacement score (Appendix E.7). Each configuration made 10 generation attempts per pocket. Table 1 reports clash rates, valid yields, and molecular-glue-related properties of the valid, clash-free molecules; evaluation details are provided in Appendix F.

TSeaDiff achieved the lowest MG-PDB-standardized PLI profile distance among the evaluated models: 0.205, compared with 0.360 for the best-performing baseline on this metric. It also attained the second-highest HDR99 inclusion rate, closely matching PocketXMol (Table 1). These results support the use of TSea for learning MG-like interaction profiles, with the clearest advantage in the agreement of normalized protein–ligand interaction patterns with the experimental reference.

PocketXMol also performed strongly, achieving the most favorable full-pocket and weaker-side Vina scores and the highest HDR inclusion rate. TSeaDiff was trained on 87,468 TSea complexes, whereas PocketXMol draws on broader multi-source training (Peng et al., 2026). Differences in training scale and coverage may contribute to its strong performance, although their effects are not isolated from differences in model design.

ODesign showed metric-dependent differences between input representations: separate-chain inputs gave higher HDR inclusion rates, whereas merged-chain inputs gave more favorable mean Vina scores and a lower PLI profile distance. Additional results in Appendix F.5 show that TSeaDiff achieved the highest diversity in the baseline comparison, but lower valid yield and QED than several baselines. Architecture ablations further showed higher valid yields and lower PLI profile distances for the full model than for both variants (Appendix F.6).

Table 1: Molecular generation results on the held-out test pockets

Model	Clash↓ (%)	Valid yield↑ (%)	Full-pocket Vina (kcal/mol)	Weaker-side Vina (kcal/mol)	PLI dist.↓	In HDR99 (%)
TSeaDiff	**3.18**	76.1	-5.23 ± 2.81	-1.19 ± 1.69	**0.205**	**95.53**
PocketXMol	**1.95**	**85.6**	-6.91 ± 2.70	-1.75 ± 1.62	0.377	**96.26**
TargetDiff	10.3	74.0	-5.57 ± 2.62	-1.40 ± 1.45	0.383	90.68
DiffSBDD	24.45	55.3	-4.58 ± 2.75	-1.01 ± 1.57	**0.360**	88.79
ODesign A/B	13.33	85.2	-4.81 ± 3.09	-0.91 ± 1.75	0.461	93.54
ODesign merged	11.36	**87.4**	-5.08 ± 2.84	-1.09 ± 1.51	0.434	92.33

6 Conclusion and Limitations

We introduced **TSea**, a large-scale synthetic structural dataset designed to address the scarcity of experimentally resolved molecular-glue ternary complexes. By partitioning binary protein–ligand complexes under structural and MG-PDB-derived contact constraints, TSea provides over 100,000 synthetic ternary environments while preserving source ligand poses and local pocket geometry. Comparative analyses show closer agreement with MG-PDB than TernaryDB on several measured properties, particularly those describing ligand engagement with the two protein partners and local pocket burial.

We further developed **TSeaDiff**, a molecular generator trained on TSea that combines pairwise protein–protein and ligand–protein interaction modeling with joint ternary denoising. On held-out, locally perturbed test pockets, TSeaDiff achieved the lowest MG-PDB-standardized protein–ligand interaction-profile distance among the evaluated models, together with competitive HDR inclusion rates. These results support the use of TSea for learning MG-like interfacial protein–ligand interaction patterns. Together, TSea and TSeaDiff provide a scalable structural resource and a generative baseline for molecular modeling at ligand-mediated protein–protein interfaces.

However, two limitations remain. **First**, TSea's ternary structures are constructed by partitioning binary protein–ligand complexes, and the resulting protein components are not guaranteed to fold independently. **Second**, the proxy metrics used to evaluate molecular-glue-like properties are based on structural, interaction, and scoring characteristics and do not directly establish molecular-glue activity in real biological systems.

Beyond molecular generation, future work could explore TSea as complementary structural supervision for ternary-complex prediction and broader molecular-glue modeling. An important direction is to investigate whether combining synthetic and experimentally resolved complexes, together with additional energetic, thermodynamic, and experimental constraints, improves generalization to real molecular-glue systems.

AI use statement

We used generative AI tools to assist with manuscript language editing and research coding, including drafting, debugging, and refining data-processing and analysis scripts. All AI-assisted text, code, and outputs were reviewed by the authors, and the resulting code and analyses were checked and verified before inclusion in the manuscript. Generative AI tools were not used to independently generate experimental data, determine the reported results, or make final scientific conclusions. Scientific interpretations and conclusions were determined by the authors based on verified results. The authors take full responsibility for the final manuscript, code, analyses, and reported findings.

Reproducibility statement

The main text and appendices describe the TSea construction pipeline, structural filtering criteria, dataset partitions, and interaction analyses, together with the TSeaDiff architecture, training objectives, and sampling procedures. We also document the locally perturbed test-set construction, baseline configurations, architecture ablations, evaluation metrics, and sampling budgets. An

anonymized repository accompanying the submission provides the code used for dataset construction and analysis, together with the dataset snapshot and metadata used to produce the reported results.

REFERENCES

Rebecca F. Alford, Andrew Leaver-Fay, Jeliazko R. Jeliazkov, Matthew J. O'Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, P. Douglas Renfrew, Vikram K. Mulligan, Kalli Kappel, Jason W. Labonte, Michael S. Pacella, Richard Bonneau, Philip Bradley, Roland L. Dunbrack, Jr., Rhiju Das, David Baker, Brian Kuhlman, Tanja Kortemme, and Jeffrey J. Gray. The Rosetta all-atom energy function for macromolecular modeling and design. *Journal of Chemical Theory and Computation*, 13(6):3031–3048, 2017. doi: 10.1021/acs.jctc.7b00125.

Guy W. Bemis and Mark A. Murcko. The properties of known drugs. 1. molecular frameworks. *Journal of Medicinal Chemistry*, 39(15):2887–2893, 1996. doi: 10.1021/jm9602928.

G. Richard Bickerton, Gaia V. Paolini, Jérémy Besnard, Sorel Muresan, and Andrew L. Hopkins. Quantifying the chemical beauty of drugs. *Nature Chemistry*, 4(2):90–98, 2012. doi: 10.1038/nchem.1243.

Martin Buttenschoen, Garrett M. Morris, and Charlotte M. Deane. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, 2024. doi: 10.1039/D3SC04185A.

Shu-Yu Chen and Martin Zacharias. What makes a good protein–protein interaction stabilizer: analysis and application of the dual-binding mechanism. *ACS Central Science*, 9(5):969–979, 2023. doi: 10.1021/acscentsci.3c00003.

Peter Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*, 1(1):8, 2009. doi: 10.1186/1758-2946-1-8.

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

Francis Gaudreault, Matthieu Chartier, and Rafael Najmanovich. Side-chain rotamer changes upon ligand binding: common, crucial, correlate with entropy and rearrange hydrogen bonding. *Bioinformatics*, 28(18):i423–i430, 2012. doi: 10.1093/bioinformatics/bts395.

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*, 2023. URL `https://openreview.net/forum?id=kJqXEPXMsE0`.

Matthew Holderfield, Bianca J. Lee, Jingjing Jiang, Aidan Tomlinson, Kyle J. Seamon, Alessia Mira, Enrico Patrucco, Grace Goodhart, Julien Dilly, Yevgeniy Gindin, Nuntana Dinglasan, Yingyun Wang, Lick Pui Lai, Shurui Cai, Lingyan Jiang, Nicole Nasholm, Nataliya Shifrin, Cristina Blaj, Harshit Shah, James W. Evans, Nilufar Montazer, Oliver Lai, Jade Shi, Ethan Ahler, Elsa Quintana, Stephanie Chang, Anthony Salvador, Abby Marquez, Jim Cregg, Yang Liu, Anthony Milin, Anqi Chen, Tamar Bar Ziv, Dylan Parsons, John E. Knox, Jennifer E. Klomp, Jennifer Roth, Matthew Rees, Melissa Ronan, Antonio Cuevas-Navarro, Feng Hu, Piro Lito, David Santamaria, Andrew J. Aguirre, Andrew M. Waters, Channing J. Der, Chiara Ambrogio, Zhengping Wang, Adrian L. Gill, Elena S. Koltun, Jacqueline A. M. Smith, David Wildes, and Mallika Singh. Concurrent inhibition of oncogenic and wild-type RAS-GTP for cancer therapy. *Nature*, 629(8013):919–926, 2024. doi: 10.1038/s41586-024-07205-6.

Rob J. Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50(2):120–126, 1996. doi: 10.1080/00031305.1996.10474359.

Rui Jiao, Xiangzhe Kong, Yinjun Jia, Yijia Zhang, Ziyi Yang, Yang Liu, and Jianzhu Ma. Scalable peptide design via memory-efficient equivariant transformer. arXiv, 2026. URL https://arxiv.org/abs/2606.25006. Preprint.

Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983. doi: 10.1002/bip.360221211.

Nickolay A. Khazanov and Heather A. Carlson. Exploring the composition of protein-ligand binding sites on a large scale. *PLOS Computational Biology*, 9(11):e1003321, 2013. doi: 10.1371/journal.pcbi.1003321.

Elizabeth A. King, Margot Meyers, and Daniel K. Nomura. Induced proximity-based therapeutic modalities. *Nature Reviews Drug Discovery*, 25(3):175–203, 2026. doi: 10.1038/s41573-025-01316-z.

Xiangzhe Kong, Junwei Chen, Ziting Zhang, Gaodeng Li, Qingyuan Zhu, Lei Wei, Mingyu Li, Yan Shi, Weiyang Dai, Zishen Zhang, Wenjuan Tan, Rui Jiao, Xiaolun Wang, Jiqing Zheng, Ziyang Yu, Qilong Wu, Zhiye Guo, Li Zhang, Wentao Li, Qiaojing Huang, Tian Zhu, Xiaowo Wang, Wenbing Huang, Yuli She, Jian Zhang, Yang Liu, Kai Liu, and Jianzhu Ma. Programming biomolecular interactions with all-atom generative model. bioRxiv, 2026. URL https://www.biorxiv.org/content/10.64898/2026.03.12.711044. Preprint.

Tanja Kortemme and David Baker. A simple physical model for binding energy hot spots in protein–protein complexes. *Proceedings of the National Academy of Sciences*, 99(22):14116–14121, 2002. doi: 10.1073/pnas.202485799.

Yiyan Liao, Jintao Zhu, Juan Xie, Luhua Lai, and Jianfeng Pei. Benchmarking cofolding methods for molecular glue ternary structure prediction. *Journal of Chemical Information and Modeling*, 65(20):11136–11148, 2025. doi: 10.1021/acs.jcim.5c01860.

Cristina Mayor-Ruiz, Sophie Bauer, Matthias Brand, Zuzanna Kozicka, Marton Siklos, Hana Imrichova, Ines H. Kaltheuner, Elisa Hahn, Kristina Seiler, Anna Koren, Georg Petzold, Michaela Fellner, Christoph Bock, André C. Müller, Johannes Zuber, Matthias Geyer, Nicolas H. Thomä, Stefan Kubicek, and Georg E. Winter. Rational discovery of molecular glue degraders via scalable chemical profiling. *Nature Chemical Biology*, 16(11):1199–1207, 2020. doi: 10.1038/s41589-020-0594-x.

Xingang Peng, Ruihan Guo, Fenglin Guo, Ziyi Wang, Jiayu Sun, Jiaqi Guan, Yinjun Jia, Yan Xu, Yanwen Huang, Muhan Zhang, Jian Peng, Xinquan Wang, Chuanhui Han, Zihua Wang, and Jianzhu Ma. Unified modeling of 3D molecular generation via atomic interactions with PocketX-Mol. *Cell*, 189(7):1904–1922.e28, 2026. doi: 10.1016/j.cell.2026.01.003.

Huan Rui, Kate S. Ashton, Jaeki Min, Connie Wang, and Patrick Ryan Potts. Protein–protein interfaces in molecular glue-induced ternary complexes: classification, characterization, and prediction. *RSC Chemical Biology*, 4(3):192–215, 2023. doi: 10.1039/d2cb00207h.

Philipp Schake, Sarah Naomi Bolz, Katja Linnemann, and Michael Schroeder. PLIP 2025: introducing protein–protein interactions to the protein–ligand interaction profiler. *Nucleic Acids Research*, 53(W1):W463–W465, 2025. doi: 10.1093/nar/gkaf361.

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.

Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.

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.

Clément Vignac, Nagham Osman, Laura Toni, and Pascal Frossard. MiDi: Mixed graph and 3D denoising diffusion for molecule generation. In *Machine Learning and Knowledge Discovery in Databases: Research Track*, volume 14170 of *Lecture Notes in Computer Science*, pp. 560–576. Springer, 2023. doi: 10.1007/978-3-031-43415-0_33.

Fanglei Xue, Meihan Zhang, Shuqi Li, Xinyu Gao, James A. Wohlschlegel, Wenbing Huang, Yi Yang, and Weixian Deng. SE(3)-equivariant ternary complex prediction towards target protein degradation. *Nature Communications*, 16(1):5514, 2025. doi: 10.1038/s41467-025-61272-5.

Jianyi Yang, Ambrish Roy, and Yang Zhang. BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions. *Nucleic Acids Research*, 41(D1):D1096–D1103, 2013. doi: 10.1093/nar/gks966.

Chengxin Zhang, Xi Zhang, Lydia Freddolino, and Yang Zhang. BioLiP2: an updated structure database for biologically relevant ligand–protein interactions. *Nucleic Acids Research*, 52(D1): D404–D412, 2024. doi: 10.1093/nar/gkad630.

Odin Zhang, Xujun Zhang, Haitao Lin, Cheng Tan, Qinghan Wang, Yuanle Mo, Qiantai Feng, Gang Du, Yuntao Yu, Zichang Jin, Ziyi You, Peicong Lin, Yijie Zhang, Yuyang Tao, Shicheng Chen, Jack Xiaoyu Chen, Chenqing Hua, Weibo Zhao, Runze Ma, Yunpeng Xia, Kejun Ying, Jun Li, Yundian Zeng, Lijun Lang, Peichen Pan, Hanqun Cao, Zihao Song, Bo Qiang, Jiaqi Wang, Pengfei Ji, Lei Bai, Jian Zhang, Chang-yu Hsieh, Pheng Ann Heng, Siqi Sun, Tingjun Hou, and Shuangjia Zheng. ODesign: a world model for biomolecular interaction design. arXiv, 2025. URL `https://arxiv.org/abs/2510.22304`. Preprint.

Jintao Zhu, Yiyan Liao, Haoyu Lin, Juan Xie, Zhichao Deng, Jinyu Han, Zhen Zhang, Jinchuan Xiao, Zhiyao Wang, Shuaipeng Zhang, Luhua Lai, and Jianfeng Pei. MGTbind: a comprehensive database of molecular glue ternary interactome. *Nucleic Acids Research*, 54(D1):D1500–D1509, 2026. doi: 10.1093/nar/gkaf1075.

A KDE-Based Contact Screening

Reference features and standardization. For each of the $n = 221$ MG-PDB reference structures, let $\mathbf{u}_i = (\overline{C}_{\mathrm{LP}}^{(i)}, \overline{N}_{\mathrm{PP}}^{(i)})^\top$ denote the two contact descriptors defined in the main text. The descriptors were standardized separately using their reference-set means and sample standard deviations:

$$\mathbf{z}_i = T(\mathbf{u}_i) = \begin{bmatrix} \left(\overline{C}_{\mathrm{LP}}^{(i)} - \mu_{\mathrm{LP}}\right)/\sigma_{\mathrm{LP}} \\ \left(\overline{N}_{\mathrm{PP}}^{(i)} - \mu_{\mathrm{PP}}\right)/\sigma_{\mathrm{PP}} \end{bmatrix}. \tag{7}$$

The reference means were $\mu_{\mathrm{LP}} = 38.369$ and $\mu_{\mathrm{PP}} = 46.240$, and the corresponding sample standard deviations were $\sigma_{\mathrm{LP}} = 17.228$ and $\sigma_{\mathrm{PP}} = 25.312$, computed using denominator $n - 1$. These statistics were held fixed throughout bandwidth selection and subsequent candidate screening. Thus, leave-one-out evaluation excludes a reference structure from the kernel sum but does not recompute the standardization parameters.

Bandwidth selection. The Gaussian KDE in Equation 4 uses the bandwidth matrix $\mathbf{H} = h^2\mathbf{I}_2$ in the standardized descriptor space. For each reference structure i, the leave-one-out (LOO) density at $\mathbf{z}_i$ was evaluated using the remaining $n - 1$ reference structures:

$$\widehat{p}_{-i,h}(\mathbf{z}_i) = \frac{1}{2\pi(n-1)h^2} \sum_{\substack{j=1 \\ j\neq i}}^{n} \exp\left(-\frac{\|\mathbf{z}_i - \mathbf{z}_j\|_2^2}{2h^2}\right). \tag{8}$$

The bandwidth was selected by maximizing the summed LOO log-likelihood:

$$\begin{aligned} \ell_{\mathrm{LOO}}(h) &= \sum_{i=1}^{n} \log \widehat{p}_{-i,h}(\mathbf{z}_i), \\ h^* &= \arg\max_{h\in\mathcal{H}} \ell_{\mathrm{LOO}}(h). \end{aligned} \tag{9}$$

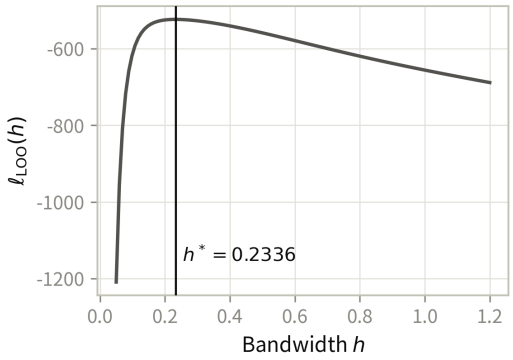


Figure 8: **Maximizing the summed LOO log-likelihood.**

The search set $\mathcal{H}$ consisted of 120 equally spaced candidate values between 0.05 and 1.2, yielding the selected bandwidth $h^* = 0.2336$ (Fig. 8). The final reference density was then fitted to all 221 structures using this bandwidth:

$$\widehat{p}_{\mathrm{ref}}(\mathbf{z}) = \widehat{p}_{h^*}(\mathbf{z}). \tag{10}$$

Each reference structure was therefore excluded from its own density evaluation during bandwidth selection and included in the final reference KDE.

Highest density regions. The 95% and 99% highest density regions (HDRs) were defined using the same fitted reference density (Hyndman, 1996). For a target probability content $q \in \{0.95, 0.99\}$, let

$$R_q = \left\{ \mathbf{z} \in \mathbb{R}^2 \,\middle|\, \widehat{p}_{\mathrm{ref}}(\mathbf{z}) \geq \tau_q \right\}, \tag{11}$$

where the density threshold τ_q satisfies

$$\int_{R_q} \widehat{p}_{\mathrm{ref}}(\mathbf{z})\, \mathrm{d}\mathbf{z} = q. \tag{12}$$

The resulting regions and thresholds satisfy

$$R_{0.95} \subseteq R_{0.99}, \qquad \tau_{0.95} \geq \tau_{0.99}. \tag{13}$$

The 95% HDR therefore provides a stricter screening criterion, whereas the 99% HDR provides a more inclusive alternative. Here, q refers to probability mass under the fitted KDE, rather than the exact fraction of observed reference structures inside the region.

Numerical construction of the HDRs. The density thresholds were determined numerically on a 240×240 uniform grid covering $[-4.5, 4.5]^2$ in the standardized descriptor space. The grid spacing in each direction was $\Delta z = 9/239$. For grid locations $\{\mathbf{g}_r\}_{r=1}^{G}$, where $G = 240^2$, the approximate probability contributions and total grid mass were

$$m_r = \widehat{p}_{\mathrm{ref}}(\mathbf{g}_r)(\Delta z)^2, \qquad M_{\mathrm{grid}} = \sum_{r=1}^{G} m_r. \tag{14}$$

The grid covered essentially all probability mass of the fitted KDE, and no renormalization was applied.

Let $d_{(1)} \geq d_{(2)} \geq \cdots \geq d_{(G)}$ denote the grid densities sorted in descending order. Their probability contributions were accumulated until the target probability was reached, giving the numerical approximation

$$k_q = \min \left\{ k : (\Delta z)^2 \sum_{r=1}^{k} d_{(r)} \geq q \right\}, \qquad \tau_q \approx d_{(k_q)}. \tag{15}$$

The resulting density thresholds were $\tau_{0.95} = 0.01704$ and $\tau_{0.99} = 0.004684$, with grid-estimated probability contents of approximately 0.9500 and 0.9900, respectively. The KDE and HDR contours in standardized space are shown in Figure 9.

Contours at these two density levels were extracted from the grid and transformed back to the original $(\overline{C}_{\mathrm{LP}}, \overline{N}_{\mathrm{PP}})$ coordinates using T^{-1}. The resulting boundaries were stored as interpolated contour polygons rather than as collections of selected grid cells. The 95% HDR has one interior hole, which is excluded from the screening region, and is nested within the 99% HDR.

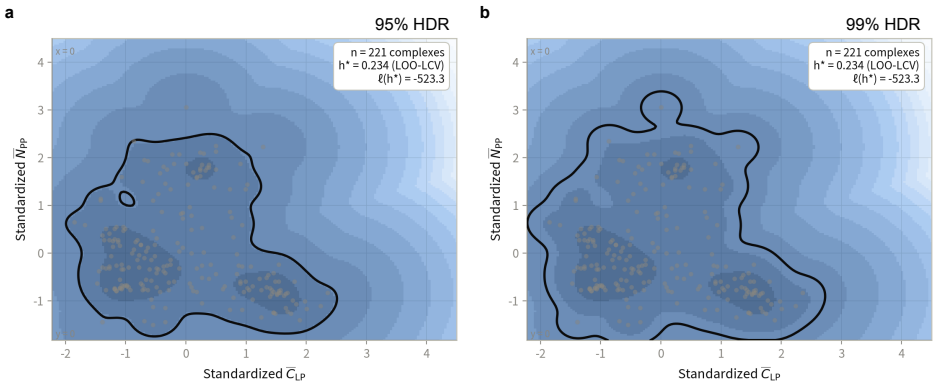


Figure 9: **2D KDE in standardized space.** Black contours indicate the **(a)** 95% and **(b)** 99% highest-density regions (HDRs).

Restriction to nonnegative contact values. Because both contact descriptors are nonnegative by definition, the physically valid domain in the original descriptor space is $\mathcal{D}_+ = [0, \infty)^2$. The corresponding restricted screening regions are

$$R_q^+ = T^{-1}(R_q) \cap \mathcal{D}_+. \tag{16}$$

The 95% contour lies entirely within this domain, whereas the 99% contour extends slightly below $\overline{N}_{\mathrm{PP}} = 0$. The extracted contour polygons were therefore clipped at the coordinate axes without refitting the KDE or changing the density thresholds.

For the fitted Gaussian KDE, the probability mass outside the nonnegative domain can be expressed as

$$\epsilon_- = 1 - \frac{1}{n} \sum_{i=1}^{n} \Phi\left(\frac{\overline{C}_{\mathrm{LP}}^{(i)}}{h^* \sigma_{\mathrm{LP}}}\right) \Phi\left(\frac{\overline{N}_{\mathrm{PP}}^{(i)}}{h^* \sigma_{\mathrm{PP}}}\right), \tag{17}$$

where Φ is the standard normal cumulative distribution function. The Gaussian-tail calculation gave an excluded probability mass of approximately 0.191%. Thus, more than 99.8% of the fitted probability mass already lies within the physically valid domain. The nominal HDR probability levels were defined before this geometric clipping; the KDE was not renormalized on $\mathcal{D}_+$.

Candidate screening and interpretation. For each candidate structure, we computed the same contact descriptors and applied the fixed transformation in Equation 7. For a candidate with nonnegative descriptor vector $\mathbf{u}_{\mathrm{cand}}$, the continuous KDE criterion at level q is

$$\widehat{p}_{\mathrm{ref}}\left(T(\mathbf{u}_{\mathrm{cand}})\right) \geq \tau_q, \tag{18}$$

equivalently, $\mathbf{u}_{\mathrm{cand}} \in R_q^+$. The stored contour polygons provide numerical representations of these screening regions. Both screening levels use the same descriptor definitions, standardization parameters, and fitted reference density; only the density threshold differs.

These criteria restrict candidate contact patterns to HDRs of the available experimental reference distribution. They constrain the admissible contact feature space without enforcing exact distributional matching between the selected candidates and MG-PDB. Region membership is a contact-based geometric screening condition and does not by itself establish ternary-complex stability or molecular glue activity.

B MG LIGAND ANALYSIS IN MG-PDB

We analyzed ligand size, contact coverage, and ligand–protein contact centrality across the 221 experimental ternary structures in MG-PDB. The distributions are shown as stacked histograms grouped by molecular glue research system.

Ligand size. Ligand heavy-atom counts ranged from 10 to 70, with most structures containing ligands of approximately 20–40 heavy atoms (Figure 10). The distribution varied across molecular glue systems. Plant-hormone ligands contributed mainly to the smaller-size region, whereas KRAS–CYPA ligands and most rapalogs occupied the upper end of the distribution, at approximately 60–70 heavy atoms. The 14-3-3 and CRBN degrader systems contributed predominantly to the intermediate-size region. These differences highlight the range of ligand sizes represented in the experimental reference. For TSea curation, we retained the observed lower bound and modestly extended the upper size cutoff to 75 heavy atoms, yielding $10 \leq |L| < 75$, to accommodate variation beyond the available MG-PDB structures.

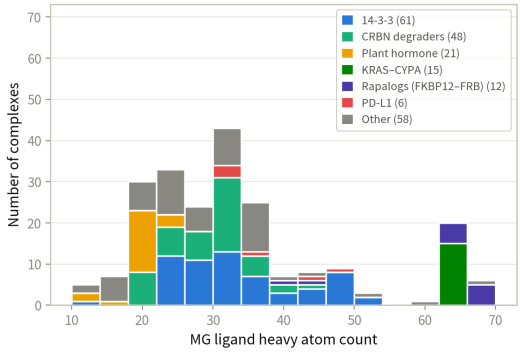


Figure 10: **Ligand size in MG-PDB.**

Ligand contact coverage. The distribution of f_{contact} was concentrated toward higher values, with most structures falling between approximately 0.6 and 1.0 (Fig. 11). CRBN degrader complexes were concentrated in the high-coverage region, whereas 14-3-3 complexes showed a broader distribution and accounted for most of the lower-coverage cases. Thus, a large fraction of ligand heavy atoms typically contacted the combined protein environment, although the extent of contact coverage differed across systems. In TSea construction, the requirement $f_{\text{contact}} > 0.4$ was applied to source binary complexes to exclude those with limited ligand–protein contact coverage as most of the existing ternary structures fall in this range.

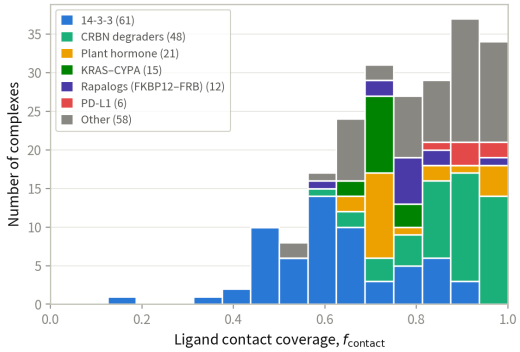


Figure 11: **Ligand contact coverage in MG-PDB.**

Ligand–protein contact centrality. In contrast to the predominantly high contact coverage, c_{LP} spanned nearly the full interval from 0 to 1 (Fig. 12). Most plant-hormone complexes and many 14-

3-3 complexes exhibited low centrality, indicating asymmetric ligand-contact counts between the two protein partners. CRBN degrader complexes were more concentrated at intermediate centrality values, whereas KRAS–CYPA and PD-L1 complexes generally exhibited higher centrality and more balanced contact counts. The experimental reference therefore includes both strongly asymmetric and comparatively balanced ligand-contact arrangements. For TSea, the requirement $c_{\mathrm{LP}} > 0.3$ limits strongly one-sided contacts in synthetic partitions while still allowing unequal engagement of the two protein components.

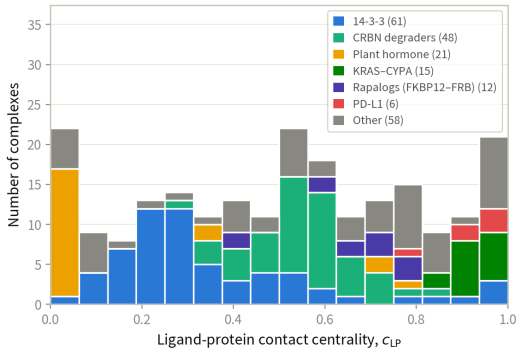


Figure 12: **Ligand–protein contact centrality in MG-PDB.**

C Detailed TSea dataset mining pipeline

C.1 Source records and coordinate normalization

We mined the *regular* small-molecule interactions in the BioLiP2 snapshot of June 29, 2026. Before reading and partitioning a structure, we excluded ligand CCD codes in a frozen 458-entry nuisance-ligand list containing common crystallization, buffer, and purification components. Let $\mathbb{P}$ and $\mathbb{L}$ denote the remaining protein and ligand heavy-atom sets. A source complex was retained only if

$$10 \le |\mathbb{L}| < 75, \qquad d_{\min}(\mathbb{P}, \mathbb{L}) \ge 2.5\ \text{Å}, \qquad f_{\text{contact}} > 0.4, \tag{19}$$

where $d_{\min}$ is the minimum atomic pair distance used to exclude potential covalent contacts and severe steric clashes.

C.2 Partition enumeration and structural screening

We assigned secondary structure using DSSP (Kabsch & Sander, 1983). For a receptor with ordered standard residues $P = (r_1, \ldots, r_n)$, every adjacent pair $i \mid i+1$ was treated as a candidate anchor. The production setting used an excision gap of two residues: r_i and r_{i+1} were removed, and the two protein components were defined as

$$A = (r_1, \ldots, r_{i-1}), \qquad B = (r_{i+2}, \ldots, r_n). \tag{20}$$

The original coordinates of all retained protein atoms and of the ligand were left unchanged. We required both A and B to contain at least 20 residues, both excised residues to have a DSSP loop code in $\{\mathrm{T}, \mathrm{S}, -, \text{blank}, \mathrm{C}\}$, and the minimum heavy-atom distance from the excised gap to L to be at least 6.0 Å. We also rejected a partition when DSSP identified a contiguous parallel or antiparallel beta ladder of at least three E–E residue pairs crossing from A to B. This last condition avoids defining two components by cutting through an intact beta sheet.

C.3 Contact screening and experimental-reference matching

For every surviving partition, we recomputed the contact descriptors defined in Section 3 after removing the two-residue gap. The production implementation used the inclusive convention $d \le 4.0$ Å. In particular, $\overline{C}_{\mathrm{LP}}$ was calculated from ligand–protein heavy-atom pair counts, whereas

$\overline{N}_{\mathrm{PP}}$ used distinct contacting protein atoms on the two sides. The component-specific counts entering c_{LP} used distinct ligand atoms; a ligand atom contacting both components contributed once to each side. We rejected partitions with undefined c_{LP} or $c_{\mathrm{LP}} \leq 0.3$, and retained only points inside the HDR99 fitted to the 221 MG-PDB reference structures.

C.4 Deduplication and materialization

Multiple nearby anchors in the same loop can produce almost identical structures. We therefore retained at most one anchor from each maximal contiguous DSSP loop, required the ordered pair $(C_{L,A}, C_{L,B})$ to be unique, and required any two retained anchor indices to differ by at least five residues.

For each retained candidate, we wrote seven coordinate views (A, B, L, AB, AL, BL, and ABL) together with a metadata record containing the source identifiers, cut definition, structural-screening results, all contact counts, HDR resource hash, and output paths. No capping, side-chain rebuilding, coordinate relaxation, or ligand redocking was performed. The resulting structures should therefore be interpreted as synthetic ternary pockets derived from experimentally observed binary pocket geometries, not as evidence that A and B are independently stable or form a native ternary assembly.

The reported HDR95 cohort was obtained by applying the frozen HDR95 membership test to the final HDR99 candidates. It is a nested label on the HDR99 release, not the result of a second partitioning, ranking, or deduplication run.

C.5 Construction algorithm

C.6 Efficient all-cut contact evaluation

Rebuilding a spatial index for every anchor would make whole-database mining unnecessarily expensive. For each source complex, the implementation builds one C-backed KD-tree over protein and ligand heavy atoms. Ligand–protein pair counts are accumulated per protein residue and converted to prefix sums, which provide $C_{L,A}$ and $C_{L,B}$ for any prefix–gap–suffix partition. For every ligand atom, the minimum and maximum contacted protein-residue indices are sufficient to obtain $N_{A\to L}$ and $N_{B\to L}$. Similarly, the minimum and maximum cross-contacting residue indices of each protein atom are combined with difference arrays to obtain $N_{A\to B}$ and $N_{B\to A}$ for all two-residue gaps in linear time after neighbor enumeration. Thus, each source record requires one contact-map construction and one DSSP assignment rather than a separate all-atom calculation for every candidate anchor.

C.7 Exact thresholds and boundary conventions

Table 2 lists the frozen production settings. The inequality direction is part of the data definition: the contact cutoff is inclusive; the upper ligand-size bound and the contact-coverage and contact-centrality thresholds are strict; the 2.5 Å source clash and 6.0 Å cut–ligand thresholds are inclusive on their accepted side.

C.8 Construction waterfall and output statistics

The frozen BioLiP2 manifest contained 501,072 regular interaction records. Table 3 summarizes the reduction in dataset size across the major source-record and partition-level stages. The 214,726 records passing source-level screening and DSSP assignment collectively contributed 79,870,535 adjacent-residue anchors. After structural screening, contact-based reference matching, and within-source deduplication, the final HDR99 release contained 106,436 synthetic ternary structures derived from 70,979 source interactions. The HDR95 cohort contains 88,277 structures and is a nested subset of the HDR99 release rather than the result of a separate construction run.

Algorithm 1 Mining the TSea HDR99 dataset from BioLiP2 binary complexes

Require: Regular BioLiP2 interactions $\mathcal{D}$
Require: Nuisance CCD set $\mathcal{B}$ and HDRs $\mathcal{H}_{99}, \mathcal{H}_{95}$
Ensure: HDR99 set $\mathcal{S}_{99}$ and nested HDR95 set $\mathcal{S}_{95}$

$\mathcal{S}_{99} \leftarrow \emptyset$
for all $(P, L, m) \in \mathcal{D}$ **do**
  if $\mathrm{CCD}(L) \in \mathcal{B}$ **then**
    continue
  end if
  $(P, L) \leftarrow \textsc{Normalize}(P, L)$
  if P is not single-chain or $|L| \notin [10, 75)$ **then**
    continue
  end if
  $M \leftarrow \textsc{ContactMap}(P, L, 4.0\ \text{Å})$
  if $d_{\min}(P, L) < 2.5$ Å or $f_{\mathrm{contact}}(M) \leq 0.4$ **then**
    continue
  end if
  $D \leftarrow \mathrm{DSSP}(P)$
  if D failed **then**
    continue
  end if
  $\mathcal{T} \leftarrow \emptyset$
  for $i = 1$ **to** $|P| - 1$ **do**
    $(A, G, B) \leftarrow \textsc{ExciseTwoResidues}(P, i)$
    if not $\textsc{StructuralCutOK}(A, G, B, L, D)$ **then**
      continue
    end if
    $(q, c_{\mathrm{LP}}) \leftarrow \textsc{ContactDescriptors}(M, A, B)$
    if not $\textsc{ContactFilterOK}(q, c_{\mathrm{LP}}, \mathcal{H}_{99})$ **then**
      continue
    end if
    $s_{\mathrm{HDR}} \leftarrow \textsc{BoundaryMargin}(q, \mathcal{H}_{99})$
    $s_{\mathrm{len}} \leftarrow 2\min(|A|, |B|)/(|A| + |B|)$
    add the annotated partition to $\mathcal{T}$
  end for
  $\mathcal{K} \leftarrow \textsc{Deduplicate}(\mathcal{T})$
  $\textsc{WriteCandidates}(\mathcal{K})$
  $\mathcal{S}_{99} \leftarrow \mathcal{S}_{99} \cup \mathcal{K}$
end for
$\mathcal{S}_{95} \leftarrow \textsc{LabelHDR95}(\mathcal{S}_{99}, \mathcal{H}_{95})$
return $\mathcal{S}_{99}, \mathcal{S}_{95}$

Here, StructuralCutOK applies the fragment-length, DSSP-loop, cross-fragment beta-ladder, and cut–ligand-distance criteria in Table 2. ContactFilterOK requires a defined $c_{\mathrm{LP}} > 0.3$ and $q = (\overline{C}_{\mathrm{LP}}, \overline{N}_{\mathrm{PP}})$ inside HDR99. Deduplicate uses the rules described in Appendix C.4.

D Dataset Evaluation

Throughout this paper, numerical results with $\pm$ are reported as mean $\pm$ standard deviation (SD), and error bars indicate one SD; in bar plots, only the error-bar arm extending away from zero is shown.

D.1 Dataset scale and ligand statistics

Complete dataset-scale and ligand-level statistics are summarized in Table 4. We report the number of constructed ternary complexes, the number of CCD ligand identifiers, ligand heavy-atom counts, and ligand–protein contact centrality c_{LP}.

Table 2: Frozen screening and selection settings for TSea.

Stage	Quantity	Acceptance rule
Source	ligand heavy atoms	$10 \leq \lvert L\rvert < 75$
Source	ligand–protein minimum distance	$d_{\min}(P, L) \geq 2.5$ Å
Source	ligand contact coverage	$f_{\text{contact}} > 0.4$
Contact	heavy-atom contact	$d \leq 4.0$ Å
Partition	excised gap	2 adjacent residues
Partition	fragment length	$\lvert A\rvert, \lvert B\rvert \geq 20$ residues
Partition	excised secondary structure	both residues are DSSP loop
Partition	cross-fragment beta ladder	fewer than 3 consecutive E–E pairs
Partition	gap–ligand minimum distance	$d_{\min}(G, L) \geq 6.0$ Å
Contact	ligand-contact centrality	$c_{\text{LP}} > 0.3$
Contact	reference support	$(\overline{C}_{\text{LP}}, \overline{N}_{\text{PP}}) \in$ HDR99
Deduplication	anchor separation	$\lvert\Delta i\rvert \geq 5$ residues

Table 3: Construction waterfall for the frozen TSea production run

Panel A: Source-record screening

Stage	Rejected	Records remaining
BioLiP2 regular input	–	501,072
Ligand identity and size filtering	57,856	443,216
Ligand–protein geometry and contact filtering	228,349	214,867
DSSP assignment	141	214,726
At least one surviving partition	143,747	70,979

Panel B: Partition screening and deduplication

Stage	Rejected	Partitions remaining
All adjacent-residue anchors	–	79,870,535
Structural cut screening	72,008,390	7,862,145
Contact balance and HDR99 screening	7,174,052	688,093
Within-source deduplication	581,657	106,436
Final HDR99 release	–	106,436
Nested HDR95 subset	–	88,277

D.2 Fixed-pose Vina analysis

Scoring protocol. We used fixed-pose AutoDock Vina to characterize ligand–protein scoring profiles in TSea, TernaryDB, and MG-PDB. For each ternary complex, we evaluated the ligand against the combined protein partners (L–AB) and against each partner separately.

Partner-specific metrics. Let E_{AL} and E_{BL} denote the Vina scores between the ligand L and protein components A and B, respectively. Because more negative Vina scores indicate more favorable interactions, we define the stronger- and weaker-partner scores as

$$E_{\text{strong}} = \min(E_{AL}, E_{BL}), \qquad E_{\text{weak}} = \max(E_{AL}, E_{BL}). \tag{21}$$

The weaker-partner score therefore corresponds to the less favorable of the two ligand–protein interactions and provides a partner-specific measure complementary to the combined L–AB score.

To quantify the balance between the two partner-specific scores, we further define Vina score centrality as

$$c_{\text{Vina}} = 1 - \frac{|E_{AL} - E_{BL}|}{|E_{AL}| + |E_{BL}|}. \tag{22}$$

Table 4: Scale and ligand analysis

Dataset	Dataset scale	Ligand count	Ligand size	Contact balance (c_{LP})
TSea (CrossDocked)	210,522	7,089	25.69 ± 7.49	0.669 ± 0.198
TSea (HDR99)	106,436	16,130	32.34 ± 15.32	0.656 ± 0.198
TSea (HDR95)	88,277	14,485	32.74 ± 15.29	0.655 ± 0.199
TernaryDB	22,760	1,615	32.54 ± 19.71	0.554 ± 0.284
MG-PDB	221	158	33.55 ± 14.25	0.496 ± 0.302

Centrality was calculated only for structures satisfying $E_{AL} < 0$ and $E_{BL} < 0$. Under this definition, c_{Vina} approaches one as the magnitudes of the two partner-specific scores become more similar and decreases as their contributions become more asymmetric.

Results. Complete statistics are reported in Table 5. TSea exhibited substantially more favorable combined and partner-specific scores than TernaryDB, with the score distributions shifted toward the MG-PDB reference. The weaker-partner scores showed particularly close agreement: the mean values were -1.90 and -1.88 kcal/mol for HDR99 and HDR95, respectively, compared with -1.85 kcal/mol for MG-PDB and -0.97 kcal/mol for TernaryDB. TSea also showed higher Vina-score centrality, reflecting more balanced score magnitudes between the two protein partners.

Table 5: Fixed-pose AutoDock Vina scores

Dataset	Full-pocket Vina (kcal/mol)	Weaker-side Vina (kcal/mol)	Stronger-side Vina (kcal/mol)	Score balance (c_{Vina})
TSea (HDR99)	-6.70 ± 2.29	-1.90 ± 1.03	-4.80 ± 1.84	0.572 ± 0.231
TSea (HDR95)	-6.65 ± 2.25	-1.88 ± 1.02	-4.77 ± 1.81	0.571 ± 0.231
TernaryDB	-4.56 ± 3.20	-0.97 ± 1.19	-3.59 ± 2.78	0.508 ± 0.289
MG-PDB	-7.91 ± 2.23	-1.85 ± 1.44	-6.06 ± 1.79	0.444 ± 0.281

D.3 PLIP INTERACTION-PROFILE ANALYSIS

Interaction extraction. We used the Protein–Ligand Interaction Profiler (PLIP) to characterize non-covalent protein–protein (PPI) and protein–ligand (PLI) interactions in TSea, TernaryDB, and MG-PDB. We analyzed five major interaction types: hydrogen bonds, hydrophobic interactions, salt bridges, π–stacking, and π–cation interactions.

Interaction-profile descriptors. For complex record i and interaction type k, let $M_{i,k}^{AB}$ denote the number of strict interchain A–B PLIP interaction events, and let $M_{i,k}^{PL}$ denote the number of protein–ligand interaction events across both protein components. The corresponding descriptors are defined as

$$x_{i,k}^{\mathrm{PPI}} = M_{i,k}^{AB}, \qquad x_{i,k}^{\mathrm{PLI}} = \frac{M_{i,k}^{PL}}{|\mathbb{L}_i|}, \tag{23}$$

where $\mathbb{L}_i$ is the ligand heavy-atom set of record i. Thus, PPI profiles describe interaction-event counts across the complete protein–protein interface, whereas PLI profiles describe ligand-size-normalized interaction densities. The PLI descriptors were normalized at the individual-record level before subsequent aggregation.

Reference-standardized profile distance. We considered the following five major interaction types:

$$\mathcal{K} = \{\text{H-bond}, \text{hydrophobic}, \text{salt bridge}, \pi\text{-stacking}, \pi\text{-cation}\}, \qquad |\mathcal{K}| = 5. \tag{24}$$

Table 6: PLIP interaction profiles

Panel A: Protein–protein interaction (PPI)

Dataset	Event count	Event density (events/residue)	Type richness	Profile distance ↓
TSea (HDR99)	30.13 ± 9.02	0.863 ± 0.101	3.059	0.897
TSea (HDR95)	28.51 ± 8.29	0.861 ± 0.103	3.010	**0.817**
TernaryDB	36.61 ± 32.82	0.843 ± 0.128	3.008	1.179
MG-PDB	17.60 ± 9.48	0.869 ± 0.214	2.661	Ref.

Event counts by interaction type

Dataset	H-bonds	Hydrophobic	Salt bridges	π-stacking	π-cation
TSea (HDR99)	10.355	18.317	0.756	0.365	0.335
TSea (HDR95)	9.714	17.430	0.713	0.341	0.310
TernaryDB	17.037	17.478	1.238	0.489	0.372
MG-PDB	9.041	7.281	0.914	0.176	0.190

Panel B: Protein–ligand interaction (PLI)

Dataset	Event count	Event density (events/heavy atom)	Type richness	Profile distance ↓
TSea (HDR99)	9.03 ± 4.01	0.359 ± 0.172	2.386	0.663
TSea (HDR95)	8.96 ± 3.91	0.355 ± 0.171	2.386	**0.636**
TernaryDB	8.21 ± 6.15	0.408 ± 0.242	1.990	1.134
MG-PDB	10.58 ± 5.36	0.337 ± 0.159	2.460	Ref.

Event counts by interaction type

Dataset	H-bonds	Hydrophobic	Salt bridges	π-stacking	π-cation
TSea (HDR99)	4.181	3.661	0.487	0.446	0.134
TSea (HDR95)	4.070	3.708	0.479	0.444	0.133
TernaryDB	4.481	2.788	0.531	0.282	0.111
MG-PDB	2.825	7.052	0.204	0.251	0.175

Let $\mu^X_{D,k}$ denote the resulting mean descriptor for dataset D and interaction type k, where $X \in \{\mathrm{PPI}, \mathrm{PLI}\}$. Let $\mu^X_{\mathrm{MG},k}$ and $\sigma^X_{\mathrm{MG},k}$ denote the corresponding mean and sample standard deviation in MG-PDB. We defined the reference-standardized profile distance as

$$d_X(D, \mathrm{MG}) = \left[\frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \left(\frac{\mu^X_{D,k} - \mu^X_{\mathrm{MG},k}}{\sigma^X_{\mathrm{MG},k}} \right)^2 \right]^{1/2}, \qquad X \in \{\mathrm{PPI}, \mathrm{PLI}\}. \tag{25}$$

Both d_{PPI} and d_{PLI} are dimensionless, with smaller values indicating closer agreement with the MG-PDB interaction profile.

Results. Complete PLIP statistics are reported in Table 6. TSea showed smaller reference-standardized distances to MG-PDB than TernaryDB for both PPI and PLI profiles. The PPI distances were 0.897 and 0.817 for HDR99 and HDR95, respectively, compared with 1.179 for TernaryDB, while the corresponding PLI distances were 0.663 and 0.636, compared with 1.134 for TernaryDB. Protein–ligand interaction density and interaction-type richness were also closer to the MG-PDB reference. At the synthetic protein–protein interface, hydrophobic interactions were more prevalent, as expected from bringing part of the hydrophobic protein interior to the newly defined interfacial surfaces through chain partitioning.

D.4 BURIED SURFACE AREA ANALYSIS

BSA definitions. We used buried surface area (BSA) as a complementary geometric measure of interface burial in TSea, TernaryDB, and MG-PDB. For each ternary complex, we separately considered the protein–protein interface and the two protein–ligand interfaces, denoted by BSA_{AB}, BSA_{AL}, and BSA_{BL}, respectively. The total buried surface area was defined as

$$\mathrm{BSA}_{\mathrm{total}} = \mathrm{BSA}_{AB} + \mathrm{BSA}_{AL} + \mathrm{BSA}_{BL}. \tag{26}$$

To characterize the two ligand-contacting interfaces independently of the arbitrary assignment of protein labels A and B, we further defined

$$\mathrm{BSA}_{\mathrm{large}} = \max\left(\mathrm{BSA}_{AL}, \mathrm{BSA}_{BL}\right), \qquad \mathrm{BSA}_{\mathrm{small}} = \min\left(\mathrm{BSA}_{AL}, \mathrm{BSA}_{BL}\right), \tag{27}$$

and their absolute difference as

$$\Delta_{\mathrm{BSA}} = \left|\mathrm{BSA}_{AL} - \mathrm{BSA}_{BL}\right|. \tag{28}$$

The combined protein–ligand BSA was defined as

$$\mathrm{BSA}_{PL} = \mathrm{BSA}_{AL} + \mathrm{BSA}_{BL}.$$

Each pairwise BSA was calculated as $\mathrm{SASA}(X) + \mathrm{SASA}(Y) - \mathrm{SASA}(XY)$.

We also quantified the balance of ligand-associated burial using BSA centrality,

$$c_{\mathrm{BSA}} = 1 - \frac{\left|\mathrm{BSA}_{AL} - \mathrm{BSA}_{BL}\right|}{\mathrm{BSA}_{AL} + \mathrm{BSA}_{BL}}. \tag{29}$$

Higher c_{BSA} therefore indicates more balanced burial of the ligand by the two protein components.

Analysis across spatial scales. We evaluated BSA in both full structures and ligand-centered pockets defined using 8, 10, and 12 Å cutoffs. A protein residue was included if any of its heavy atoms lay within the specified cutoff of any ligand heavy atom. Per-atom SASA losses were calculated using the full components in each pairwise system and then summed over all heavy atoms of the selected protein residues. For each protein–ligand interface, the complete ligand-side SASA loss was also included. Local total BSA therefore comprises the local protein–protein BSA and both protein–ligand BSAs.

Results. Full-structure BSA components and protein–ligand interface balance are reported in Panels A and B of Table 7, respectively, while local total BSA across ligand-centered pocket scales is reported in Panel C. Across the 8, 10, and 12 Å pocket scales, TSea showed local total BSA distributions closer to MG-PDB than TernaryDB. The closest agreement in mean local total BSA occurred at 8 Å, with the deviation increasing at larger radii. At the full-structure scale, the mean BSA of the smaller protein–ligand interface remained close to MG-PDB, whereas protein–protein burial was greater.

D.5 CONSTRUCTION ABLATION

To assess whether HDR screening improves agreement with the experimental MG distribution beyond the contact descriptors used to define the screening criterion, we evaluated the resulting datasets using PLIP, BSA, and Vina-based descriptors that were not used during HDR construction. We performed the ablation on a random subset of 10,000 source PDB entries. From the same source pool, the no-HDR, HDR99, and HDR95 pipelines yielded 15,991, 8,462, and 6,946 accepted candidates, respectively. The results are summarized in Table 8.

E TSEADIFF

E.1 PROBLEM FORMULATION AND INPUT REPRESENTATION

Let $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{L}$ denote the heavy-atom sets of the two protein pockets and the ligand, respectively, and let $\mathbb{V} = \mathbb{A} \cup \mathbb{B} \cup \mathbb{L}$. Given fixed pocket conditions $\mathcal{C}_A$ and $\mathcal{C}_B$ and a ligand heavy-atom count $N_L = |\mathbb{L}|$, the model learns

$$p_\theta(\mathbf{v}_L, \mathbf{x}_L \mid \mathcal{C}_A, \mathcal{C}_B, N_L), \tag{30}$$

Table 7: Buried surface area statistics

Panel A: Full-structure BSA components			
Dataset	$\mathrm{BSA}_{\mathrm{total}}$ (Å^2)	BSA_{AB} (Å^2)	BSA_{PL} (Å^2)
TSea (HDR99)	3540 ± 932	2632 ± 826	908 ± 291
TSea (HDR95)	3385 ± 834	2479 ± 727	907 ± 285
TernaryDB	3186 ± 2719	2430 ± 2632	756 ± 397
MG-PDB	2659 ± 1068	1640 ± 1053	1019 ± 328

Panel B: Protein–ligand interfaces and balance				
Dataset	$\mathrm{BSA}_{\mathrm{large}}$ (Å^2)	$\mathrm{BSA}_{\mathrm{small}}$ (Å^2)	Δ_{BSA} (Å^2)	c_{BSA}
TSea (HDR99)	603 ± 205	305 ± 139	297 ± 194	0.669 ± 0.189
TSea (HDR95)	602 ± 201	304 ± 137	298 ± 193	0.667 ± 0.190
TernaryDB	529 ± 301	228 ± 160	301 ± 274	0.618 ± 0.249
MG-PDB	725 ± 189	294 ± 214	431 ± 236	0.535 ± 0.266

Panel C: Local total BSA across pocket scales			
Dataset	8 Å pocket (Å^2)	10 Å pocket (Å^2)	12 Å pocket (Å^2)
TSea (HDR99)	1753 ± 548	2041 ± 624	2315 ± 696
TSea (HDR95)	1732 ± 531	2007 ± 602	2264 ± 665
TernaryDB	1247 ± 670	1439 ± 817	1621 ± 970
MG-PDB	1727 ± 446	1899 ± 470	2067 ± 520

Table 8: Construction ablation results

Dataset	PPI profile dist. ↓	PLI profile dist. ↓	Protein–protein BSA (Å^2)	Weaker-side Vina (kcal/mol)	Score balance (c_{Vina})
No-HDR	2.174	0.779	4368 ± 2317	-1.91 ± 0.94	0.559 ± 0.189
HDR99	0.898	0.659	2604 ± 712	-1.84 ± 0.96	0.553 ± 0.205
HDR95	**0.807**	**0.632**	2464 ± 646	-1.82 ± 0.94	0.554 ± 0.206

where $\mathbf{v}_L$ and $\mathbf{x}_L$ denote ligand atom types and coordinates. We use seven ligand atom classes: C, N, O, F, P, S, and Cl ($K = 7$). Protein coordinates and atom types remain fixed during diffusion, while ligand bonds and bond orders are reconstructed after sampling.

Coordinate normalization. Coordinates are centered at the midpoint between the centroids of the two protein pockets, $\mathbf{c} = (\mathbf{c}_A + \mathbf{c}_B)/2$, and scaled by $s_x = 4.0$ Å. This centering is invariant to exchanging proteins A and B. All graph radii are specified in physical units and converted accordingly during graph construction.

Node features. Each node is represented by

$$\mathbf{z}_i = \mathrm{concat}\big(\mathrm{onehot}(v_{i,t}), \mathbf{e}_t(\tau), \mathbf{e}_r(r_i), \mathbf{e}_a(a_i), \mathbf{e}_n(n_i), \mathbf{e}_b(b_i), \mathbf{e}_N(N_L)\big), \tag{31}$$

where $\tau = t/T$, and the remaining terms encode entity role, residue identity, protein atom-name type, backbone status, and target ligand size.

Typed geometric graphs. Each stage uses a directed radius/KNN graph with self-loops and seven relation types: self, ligand–ligand, within-A, within-B, A–B, ligand–A, and ligand–B. To preserve

partner-exchange symmetry, within-B and ligand–B share relation embeddings with within-A and ligand–A, respectively.

Neighbors are selected within relation-specific radii and capped at the values in Table 9; selected cross-entity relations additionally use a larger-radius nearest-neighbor fallback. Expert graphs are constructed from the current noisy coordinates, and the common-stage graph is rebuilt after expert fusion.

Table 9: Geometric graph settings

Stage	Relation	Radius (Å)	Max. neighbors	Fallback
PPI	Protein intra	8	24	None
PPI	A–B	8	24	4 within 12 Å
Binding	Ligand intra	6	12	4 within 8 Å
Binding	Protein intra	6	24	None
Binding	Ligand–protein	8	24	4 within 12 Å
Common	Ligand intra	6	12	4 within 8 Å
Common	Protein intra	6	24	None
Common	A–B	6	24	4 within 12 Å
Common	Ligand–protein	8	24	4 within 12 Å

E.2 Hierarchical equivariant denoising network

The network first processes pairwise subgraphs with specialized experts and then integrates their representations on the full ternary graph. All experts and the common block use the same scalar/vector equivariant message-passing backbone, MEET (Jiao et al., 2026), with stage-specific hidden dimensions, depths, and coordinate-update masks.

Pairwise experts. The PPI expert operates on $\mathbb{A} \cup \mathbb{B}$, while a shared binding expert is applied independently to $\mathbb{L}\cup\mathbb{A}$ and $\mathbb{L}\cup\mathbb{B}$. Each expert contains one 256-dimensional MEET layer, connected to the shared 512-dimensional representation through input and output adapters. Only nodes in the corresponding pairwise subgraph participate in each expert.

The PPI expert updates protein features without modifying coordinates. The binding expert updates features of both ligand and protein nodes but permits coordinate updates only for ligand atoms. The two binding branches share all parameters.

Symmetric residual fusion. Let $\mathbf{h}_i$ denote the shared input representation and $\mathbf{h}_i^P$, $\mathbf{h}_i^A$, and $\mathbf{h}_i^B$ the corresponding expert outputs, with $\Delta\mathbf{h}_i^e = \mathbf{h}_i^e - \mathbf{h}_i$. The expert residuals are fused symmetrically as

$$\mathbf{h}_i^{\text{fuse}} = \mathbf{h}_i + \begin{cases} \frac{1}{2}(\Delta\mathbf{h}_i^A + \Delta\mathbf{h}_i^B), & i \in \mathbb{L}, \\ \frac{1}{2}(\Delta\mathbf{h}_i^P + \Delta\mathbf{h}_i^A), & i \in \mathbb{A}, \\ \frac{1}{2}(\Delta\mathbf{h}_i^P + \Delta\mathbf{h}_i^B), & i \in \mathbb{B}. \end{cases} \tag{32}$$

For ligand atoms, coordinate residuals from the two binding branches are averaged in the same manner:

$$\widetilde{\mathbf{x}}_i^{\text{fuse}} = \widetilde{\mathbf{x}}_{i,t} + \frac{1}{2}\left[(\widetilde{\mathbf{x}}_i^A - \widetilde{\mathbf{x}}_{i,t}) + (\widetilde{\mathbf{x}}_i^B - \widetilde{\mathbf{x}}_{i,t})\right], \qquad i \in \mathbb{L}. \tag{33}$$

Protein coordinates remain fixed throughout expert fusion.

Common block and output heads. The fused ternary graph is then processed by an eight-layer MEET block with hidden width 512. All MEET blocks use eight attention heads, dropout 0.1, and typed edge features. The common block updates all node features, while coordinate updates are retained only for ligand atoms.

The atom head is a $512 \to 7$ linear projection, and the final vector representation is projected to a three-dimensional coordinate residual. For ligand atoms, the network outputs

$$\boldsymbol{\ell}_{\theta,i} = \mathbf{W}_v \mathbf{h}_i^{\text{common}} + \mathbf{b}_v, \qquad \boldsymbol{\epsilon}_{\theta,i} = \widetilde{\mathbf{x}}_i^{\text{common}} - \widetilde{\mathbf{x}}_{i,t}, \qquad i \in \mathbb{L}. \tag{34}$$

The two terms represent clean atom-type logits and coordinate-noise predictions, respectively; outputs on protein nodes are masked out.

Symmetry. Shared protein entity embeddings, shared binding-expert parameters, tied relation embeddings, symmetric centering, and symmetric residual fusion make the ligand prediction invariant to exchanging proteins A and B. The scalar/vector MEET backbone further yields invariant atom-type predictions and equivariant coordinate predictions under global rigid-body transformations.

E.3 Joint atom and coordinate diffusion

Following the mixed continuous–discrete diffusion formulation of MiDi (Vignac et al., 2023), we jointly diffuse ligand coordinates with Gaussian noise and ligand atom types with a categorical process. Both modalities use $T = 800$ steps and share the normalized time $\tau = t/T$, while using separate noise schedules.

Noise schedules. For modality $m \in \{v, x\}$, we define

$$A_m(\tau) = \frac{\cos^2\left[\frac{\pi}{2}\left(\frac{\tau+s}{1+s}\right)^{p_m}\right]}{\cos^2\left[\frac{\pi}{2}\left(\frac{s}{1+s}\right)^{p_m}\right]}, \tag{35}$$

with $s = 0.01$, $p_v = 1.0$, and $p_x = 2.5$. The corresponding diffusion coefficients are

$$\beta_t^m = 1 - \frac{A_m(t/T)}{A_m((t-1)/T)}, \qquad \alpha_t^m = 1 - \beta_t^m, \qquad \overline{\alpha}_t^m = \prod_{j=1}^{t} \alpha_j^m. \tag{36}$$

We cap β_t^x at 0.30 and β_t^v at 0.999. The coordinate schedule ($p_x = 2.5$) retains geometric signal longer than the atom-type schedule, while both modalities are denoised jointly at the same timestep.

Coordinate diffusion. For normalized ligand coordinates, the forward process is

$$\widetilde{\mathbf{x}}_{L,t} = \sqrt{\overline{\alpha}_t^x}\,\widetilde{\mathbf{x}}_{L,0} + \sqrt{1-\overline{\alpha}_t^x}\,\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{37}$$

The network predicts $\boldsymbol{\epsilon}_\theta$ from the joint noisy ligand state and pocket conditions. Reverse sampling follows the standard DDPM update

$$\boldsymbol{\mu}_\theta = \frac{1}{\sqrt{\alpha_t^x}}\left(\widetilde{\mathbf{x}}_{L,t} - \frac{\beta_t^x}{\sqrt{1-\overline{\alpha}_t^x}}\boldsymbol{\epsilon}_\theta\right), \tag{38}$$

$$\widetilde{\mathbf{x}}_{L,t-1} = \boldsymbol{\mu}_\theta + \sigma_t \mathbf{z}, \qquad \sigma_t^2 = \beta_t^x \frac{1-\overline{\alpha}_{t-1}^x}{1-\overline{\alpha}_t^x}, \tag{39}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ for $t > 1$ and $\mathbf{z} = \mathbf{0}$ at the final step.

Categorical atom diffusion. Let $\boldsymbol{\pi}$ denote the uniform distribution over the $K = 7$ atom classes. The categorical forward transition is

$$\mathbf{Q}_t = \alpha_t^v \mathbf{I} + (1-\alpha_t^v)\mathbf{1}\boldsymbol{\pi}^\top, \tag{40}$$

giving

$$q(v_t = k \mid v_0 = j) = \overline{\alpha}_t^v \mathbf{1}[k=j] + (1-\overline{\alpha}_t^v)\pi_k. \tag{41}$$

The atom head predicts the clean distribution $p_\theta^0 = \mathrm{softmax}(\boldsymbol{\ell}_\theta)$. Let

$$r_{\theta,t-1}(k) = \overline{\alpha}_{t-1}^v p_\theta^0(k) + (1-\overline{\alpha}_{t-1}^v)\pi_k. \tag{42}$$

For the observed noisy atom $v_t = y$, the reverse distribution is

$$p_\theta(v_{t-1} = k \mid v_t = y, \widetilde{\mathbf{x}}_{L,t}, \mathcal{C}) = \frac{r_{\theta,t-1}(k)\, Q_t(k,y)}{\sum_{c=1}^{K} r_{\theta,t-1}(c)\, Q_t(c,y)}, \tag{43}$$

where $\mathcal{C} = (\mathcal{C}_A, \mathcal{C}_B, N_L)$. The same posterior construction is used during training and sampling.

E.4 TRAINING OBJECTIVE

The training objective combines coordinate denoising, atom-type prediction, and geometric regularization:

$$\begin{aligned}\mathcal{L} = \lambda_x \mathcal{L}_\epsilon + \lambda_{x_0} \mathcal{L}_{x_0} + \lambda_{\mathrm{KL}} \mathcal{L}_{\mathrm{atom,KL}} \\ + \lambda_{\mathrm{NLL}} \mathcal{L}_{\mathrm{atom,NLL}} + \lambda_{\mathrm{CE}} \mathcal{L}_{\mathrm{atom,CE}} + \mathcal{L}_{\mathrm{geometry}}.\end{aligned} \tag{44}$$

Coordinate objectives. The primary coordinate objective is mean-squared noise regression over ligand atoms:

$$\mathcal{L}_\epsilon = \frac{1}{\sum_{b\in\mathcal{B}} N_b} \sum_{b\in\mathcal{B}} \sum_{i\in\mathbb{L}_b} \|\boldsymbol{\epsilon}_{b,i} - \boldsymbol{\epsilon}_{\theta,b,i}\|_2^2 . \tag{45}$$

Here, $\mathcal{B}$ denotes a minibatch, $\mathbb{L}_b$ is the ligand heavy-atom set of sample b, and $N_b = |\mathbb{L}_b|$. An auxiliary $\mathcal{L}_{x_0}$ supervises the reconstructed clean ligand coordinates in normalized units and is applied only in the low-noise regime

$$\tau \le 0.5, \qquad \mathrm{SNR}_x(t) \ge 0.25. \tag{46}$$

Atom objectives. For $t > 1$, atom-type denoising is trained with the variational term

$$\mathcal{L}_{\mathrm{atom,KL}} = D_{\mathrm{KL}}(q(v_{t-1} \mid v_t, v_0) \,\|\, p_\theta(v_{t-1} \mid v_t, \widetilde{\mathbf{x}}_{L,t}, \mathcal{C})) . \tag{47}$$

At $t = 1$, it is replaced by the decoder negative log-likelihood

$$\mathcal{L}_{\mathrm{atom,NLL}} = -\log p_\theta(v_0 = v_0^{\mathrm{true}} \mid v_1, \widetilde{\mathbf{x}}_{L,1}, \mathcal{C}), \tag{48}$$

using the same categorical reverse posterior as in sampling. We additionally apply a class-weighted cross-entropy $\mathcal{L}_{\mathrm{atom,CE}}$ to the predicted clean atom distribution, with inverse-frequency weights capped at 5.

Geometric regularization. Geometric losses are evaluated on reconstructed clean ligand coordinates after restoring physical units with s_x and are scaled by $\sqrt{\overline{\alpha}_t^x}$. Bond-length, bond-angle, and dihedral terms use the low-noise gate in Equation 46. Relative radius of gyration and intra- and intermolecular clash terms require only $\mathrm{SNR}_x(t) \ge 0.25$.

E.5 SAMPLING AND IMPLEMENTATION DETAILS

Reverse sampling. Algorithm 2 summarizes the sampling procedure. Protein atom types and coordinates remain fixed throughout diffusion. Ligand bonds and bond orders are reconstructed after sampling.

Algorithm 2 TSeaDiff sampling

Input: Protein pockets $\mathcal{C}_A$ and $\mathcal{C}_B$; ligand heavy-atom count N_L; trained network f_θ; diffusion steps T
Output: Ligand atom types $\mathbf{v}_{L,0}$ and coordinates $\mathbf{x}_{L,0}$

```
Center and scale the pocket coordinates as described in Section E.1
Initialize N_L ligand atom types v_{L,T} independently from the uniform seven-class distribution
Initialize normalized ligand coordinates x̃_{L,T} ~ N(0, I)
for t = T, ..., 1 do
    Predict clean atom-type probabilities p_θ^0 and coordinate noise ε_θ from the current noisy ligand and fixed pockets using f_θ (Section E.2)
    Sample atom types v_{L,t-1} using Equation 43
    Update coordinates x̃_{L,t-1} using Equation 39, with no added noise at t = 1
end for
Restore ligand coordinates: x_{L,0} = s_x x̃_{L,0} + c
return (v_{L,0}, x_{L,0})
```

Training setup. Training uses paired timestep sampling: half of each minibatch is sampled uniformly from $\{1, \ldots, T\}$, and the remaining samples use complementary timesteps $t' = T + 1 - t$. After excluding HDR99 complexes whose ligand or input protein pockets contain elements outside the seven-element vocabulary (C, N, O, F, P, S, Cl), 89,468 complexes remain. These are partitioned into 87,468 training, 1,000 validation, and 1,000 test samples. Complexes sharing a Bemis–Murcko

scaffold (Bemis & Murcko, 1996), ligand identity, source PDB entry, or parent complex are grouped into the same partition to avoid overlap in these identifiers across partitions. The full model and all architecture ablations use the same split.

E.6 Architecture ablations

We examine the pairwise-to-ternary design using two separately trained variants, *common-only* and *joint-expert*, both of which retain access to the two protein partners. Their results are reported in Appendix F.6.

Common-only. This variant removes the entire pairwise expert stage, including the PPI expert, the two shared-parameter binding branches, and their residual fusion. The initial node representations and noisy ligand coordinates are passed directly to the eight-layer common block, which jointly processes both protein pockets and the ligand without an explicit preliminary decomposition into pairwise views.

Joint-expert. This variant replaces the pairwise expert stage with a single 256-dimensional, two-layer expert operating on the complete protein–ligand–protein system. The expert updates all node features and only ligand coordinates, using the same graph specification as the common block. Its outputs are passed to the unchanged eight-layer common block, whose graph is rebuilt from the updated coordinates. There are no separate pairwise branches or cross-view residual averaging. The name refers to a joint ternary expert, not conditioning on a single protein partner.

Comparison settings. All three configurations use identical training, validation, and test partitions, together with the same atom vocabulary, coordinate normalization, diffusion schedules, loss weights, and optimizer settings. Each configuration contains one common MEET block with eight layers of width 512. The full model additionally uses a one-layer PPI expert and a one-layer binding expert, both of width 256; the binding expert is applied twice with shared parameters, once for each ligand–protein view. The common-only variant omits the expert stage, whereas the joint-expert variant uses one two-layer expert of width 256 over the complete ternary system.

The full model, common-only variant, and joint-expert variant contain approximately 106.302, 99.337, and 105.904 million parameters, respectively. The common-only comparison examines the contribution of the additional expert stage, including its capacity. The joint-expert comparison examines explicit pairwise decomposition and symmetric fusion relative to a joint ternary expert stage with a similar parameter count. Neither comparison individually isolates the contribution of the PPI expert, parameter sharing, or residual averaging.

E.7 Test set construction

Source selection and pocket definition. We constructed the evaluation set from the 1,000 held-out test samples of TSea (HDR99), using the scaffold- and source-group-based split described above. Because chain partitioning preserves the original ligand-bound pocket geometry, we introduced local side-chain perturbations to avoid directly presenting the exact source pocket conformations to the evaluated models. Such perturbations are biologically motivated by local conformational adaptation and side-chain rearrangements associated with ligand binding (Gaudreault et al., 2012).

The perturbation pocket comprises all protein residues with at least one heavy atom within 12 Å of any reference-ligand heavy atom. We denote by $\mathbb{H}$ the set of all side-chain heavy atoms belonging to these residues. Pocket membership and distance-based weights were determined from the reference structure and held fixed across candidate conformations.

Energy-based filtering. For each prepared source complex, 1,000 candidate conformations were generated by local side-chain repacking followed by side-chain minimization in Rosetta (Alford et al., 2017). The reference protein underwent the same restricted side-chain minimization. The reference ligand defined the pocket and displacement weights but was not included in the Rosetta energy. To retain only energy-similar perturbations, candidate conformation k was required to satisfy

$$\frac{|E_k - E_0|}{N_{\mathrm{rot}}} \leq 0.5 \text{ REU/residue}, \tag{49}$$

where E_0 and E_k denote the Rosetta scores of the reference structure and minimized candidate k, respectively, evaluated using the same scoring function and molecular system. N_{rot} is the number of pocket residues with rotameric side-chain degrees of freedom; Gly and Ala are excluded from this normalization because they do not possess side-chain χ degrees of freedom. REU denotes Rosetta energy units.

Side-chain displacement scoring. Among candidates satisfying Equation 49, we favored conformations with larger side-chain rearrangements near the reference ligand. For candidate conformation k, the ligand-proximity-weighted side-chain displacement score was defined as

$$S_k = \sum_{a \in \mathbb{H}} \frac{\left\| \mathbf{r}^{(k)}_{\phi_k(a)} - \mathbf{r}^{(0)}_a \right\|_2}{\left[\min_{l \in \mathbb{L}} \left\| \mathbf{r}^{(0)}_a - \mathbf{r}^{(0)}_l \right\|_2 \right]^2}. \tag{50}$$

Here, k indexes candidate conformations generated from the same source complex. For pocket side-chain atom a, $\mathbf{r}^{(0)}_a$ denotes its coordinate in the reference structure, whereas $\mathbf{r}^{(k)}_{\phi_k(a)}$ denotes the coordinate of its corresponding atom in candidate k after symmetry correction. The numerator measures the Euclidean displacement of each pocket side-chain heavy atom.

The denominator is the squared minimum Euclidean distance between atom a and any reference-ligand heavy atom. These distances are calculated from the reference structure and held fixed across all candidate conformations from the same source complex. Consequently, side-chain atoms closer to the reference ligand receive larger weights, emphasizing rearrangements in the immediate ligand environment.

Symmetry correction. Chemically equivalent side-chain atoms can exchange atom labels without representing an actual conformational change. To prevent such relabeling from contributing to S_k, symmetry-equivalent atom mappings were considered for ASP, GLU, PHE, TYR, ARG, VAL, and LEU. For these residues, the identity mapping and the corresponding chemically equivalent permutation were evaluated, whereas only the identity mapping was used for other residue types. A single mapping was selected consistently for all relevant side-chain atoms within each residue, using the mapping that minimized the weighted displacement of that residue.

Final test-set selection. For each source complex, the energy-qualified candidate with the largest S_k was retained. The displacement score of this selected conformation is denoted by S. For cases sharing the same full CCD identifier, only the case with the largest displacement score was retained. The remaining conformations were ranked by S, and the 100 highest-scoring conformations, each originating from a distinct source complex, were retained as the test set for model comparisons and architecture ablations.

F Model Evaluation

F.1 Compared models and input preparation

We compare TSeaDiff with five external baseline configurations: PocketXMol (Peng et al., 2026), TargetDiff (Guan et al., 2023), DiffSBDD (Schneuing et al., 2024), and ODesign (Zhang et al., 2025) with separate-chain and merged-chain protein inputs. The external models use released checkpoints without fine-tuning on TSea. PocketXMol uses its small-molecule generation task, DiffSBDD uses the CrossDocked full-atom conditional checkpoint, and ODesign uses its ligand-rigid checkpoint. AnewOmni (Kong et al., 2026) was not included in the quantitative comparison because reliably matching its generated ligand heavy-atom count to that of the reference ligand was not straightforward in our evaluation setup. The architecture-ablation variants are defined in Appendix E.6, with their results reported in Appendix F.6.

All configurations use the same selected protein structures and the reference-ligand heavy-atom count. The reference ligand specifies the binding-site location and target molecular size, but its chemical graph is not provided as a scaffold for generation. Model-specific protein input preparation is retained. TSeaDiff uses complete residues with a heavy atom within 8 Å of a reference-ligand

heavy atom. PocketXMol and TargetDiff use their native 10 Å residue-center-of-mass selection rule. DiffSBDD uses standard protein residues with a heavy atom within 8 Å of the reference ligand.

ODesign receives the complete reference protein components, with hotspot residues defined by a 4.5 Å distance to the reference ligand and the reference-ligand centroid used as the generation center. The separate configuration represents A and B as two protein-chain inputs, whereas the merged configuration encodes the same components as one protein-chain input. Both configurations use the same weights and protein coordinates. All downstream interaction metrics retain the original A/B partition.

F.2 Evaluation subsets and sampling budgets

We use two complementary evaluation panels. Molecular-glue properties are evaluated on all 100 selected test pockets, with 10 generation attempts per pocket. Molecular generation quality is evaluated on a fixed subset of 10 pockets, with 100 attempts per pocket. Each panel therefore contains 1,000 attempted draws per configuration, but measures a different aspect of generation.

The same pocket lists are used for every configuration. The 100-pocket panel uses the first 10 chronological attempts for each pocket. The 10-pocket panel uses 100 total attempts per pocket, rather than 100 successful molecules. When the first 10 attempts are reused, the additional 90 attempts complete the larger sampling budget. Failed generation, reconstruction, or subsequent filtering does not trigger replacement by later samples. The two panels are summarized separately.

F.3 Validity and steric compatibility

A chemically valid output must be readable and sanitizable by RDKit, contain at least one heavy atom, form a single connected component, and have one conformer with finite coordinates. Protein clashes are evaluated using the same PoseBusters configuration on a common 12 Å reference-ligand-centered protein pocket (Buttenschoen et al., 2024). A molecule is classified as clashing if either the minimum distance to protein test or the volume overlap with protein test fails. The distance test uses a van der Waals radius scale of 0.8 and a clash cutoff of 0.75; the volume test uses a van der Waals scale of 0.8 and an overlap cutoff of 0.075. This definition does not represent the failure rate of the complete PoseBusters test suite.

Throughout the reported tables, *valid molecules* denotes chemically valid, connected outputs with no detected protein clash. Outputs with missing PoseBusters determinations are excluded rather than classified as clash-free. Valid yield is the number of retained molecules divided by the 1,000 attempts in the corresponding evaluation panel. Unless stated otherwise, subsequent metrics are computed only over these retained molecules.

F.4 Molecular quality and molecular-glue evaluation

For molecular quality, we report QED (Bickerton et al., 2012), synthetic accessibility (SA; lower is better) (Ertl & Schuffenhauer, 2009), molecular diversity, and uniqueness, together with the rate of valid molecules. These statistics are computed over valid, clash-free outputs from the 10-pocket evaluation panel.

For molecular-glue-related properties, we report full-pocket and weaker-side fixed-pose Vina scores, the MG-PDB-standardized PLI profile distance, and the fractions of valid molecules inside the HDR99 contact regions. We retain the metric definitions and experimental-reference statistics introduced in the dataset analysis. Generated ligand coordinates are evaluated at their sampled poses, without redocking or further geometry optimization.

For PLIP-based comparisons, ligand-size-normalized interaction counts are first averaged within each pocket and then equally across evaluable pockets before computing the aggregate profile distance. This distance measures agreement between mean interaction-type profiles, rather than between complete multivariate distributions. For contact-space evaluation, the MG-PDB-derived HDR boundaries are kept fixed and are not refitted to generated samples. The reported HDR percentages use valid, clash-free molecules as the denominator, rather than all generation attempts. Scoring or PLIP failures are omitted from the corresponding summaries rather than assigned zero.

These metrics characterize molecular quality and MG-like structural compatibility, but do not directly measure ternary binding cooperativity or biological activity.

F.5 Additional generation results

Table 10 summarizes molecular quality on the fixed 10-pocket subset, using 100 attempts per pocket. The 10-pocket subset was selected deterministically from the 100 test pockets using only reference structures and input features. We divided pockets into ten equal-sized bins by reference-ligand heavy-atom count and selected one from each bin using greedy max–min diversity over ligand fingerprints, input descriptors, and ligand element sets.

Table 10: Molecular quality of generated ligands.

Model	Clash↓ (%)	Valid yield↑ (%)	QED↑	SA↓	Diversity↑	Uniqueness↑ (%)
TSeaDiff	**14.63**	66.5	0.313 ± 0.165	5.02 ± 1.21	**0.681** ± 0.098	99.36
PocketXMol	**9.64**	76.9	**0.433** ± 0.185	**3.31** ± 1.27	0.621 ± 0.116	98.30
TargetDiff	23.9	60.5	**0.431** ± 0.195	5.19 ± 1.02	0.652 ± 0.125	**100.00**
DiffSBDD	53.95	32.1	0.400 ± 0.191	4.87 ± 0.93	**0.674** ± 0.128	**100.00**
ODesign A/B	20.39	**77.7**	0.376 ± 0.179	**4.37** ± 1.63	0.582 ± 0.135	88.44
ODesign merged	17.26	**81.0**	0.397 ± 0.198	4.40 ± 1.61	0.630 ± 0.127	93.98

Molecular quality. TSeaDiff achieved the highest mean diversity (0.681 ± 0.098) and near-complete uniqueness (99.36%), indicating diverse generation among the retained molecules. However, its valid yield and QED were lower than those of several baselines, and its SA score was less favorable than most. PocketXMol achieved the highest mean QED and lowest mean SA. Thus, the favorable PLI profile agreement of TSeaDiff does not imply uniformly better molecular quality. Three example molecules generated by TSeaDiff are shown in Figure 13.

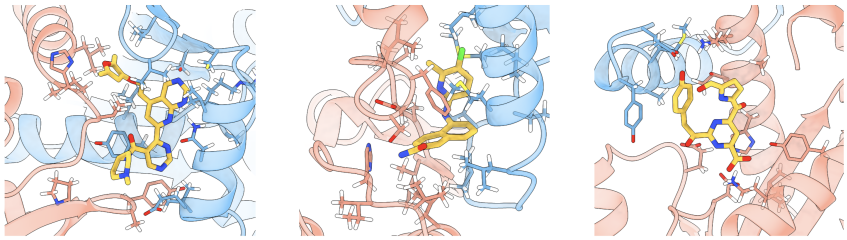


Figure 13: **Example molecules (yellow) generated by TSeaDiff.**

ODesign input representation. The two ODesign configurations used identical model weights and protein coordinates, differing in whether the protein components were supplied as separate chains or as one merged chain; all interaction metrics retained the original A/B partition. Separate-chain inputs gave higher HDR99 inclusion rates, whereas merged-chain inputs gave more favorable mean full-pocket and weaker-side Vina scores and a lower PLI profile distance (Table 1). On the molecular-quality panel, merged inputs also yielded more valid molecules, higher QED, greater diversity, and higher uniqueness, while SA changed little (Table 10). These metric-dependent differences are consistent with sensitivity to protein-chain representation in this evaluation, rather than a uniform advantage of either input encoding.

F.6 Architecture ablation results

Table 11 compares the full TSeaDiff model with the *common-only* and *joint-expert* variants defined in Appendix E.6. Panel A of Table 11 uses the 100-pocket molecular-glue evaluation, whereas Panel B uses the 10-pocket molecular-quality evaluation, with the same subsets and sampling budgets as the corresponding baseline comparisons.

Table 11: Ablation study of TSeaDiff

Panel A: Molecular-glue properties

Model	Clash↓ (%)	Valid yield↑ (%)	Full-pocket Vina (kcal/mol)	Weaker-side Vina (kcal/mol)	PLI dist.↓	In HDR99 (%)
TSeaDiff	3.18	**76.1**	**-5.23** $\pm$ 2.81	**-1.19** $\pm$ 1.69	**0.205**	**95.53**
Common-only	**2.44**	67.9	-5.06 ± 2.74	-1.10 ± 1.62	0.356	95.43
Joint-expert	3.62	71.8	-4.93 ± 2.67	-1.13 ± 1.59	0.236	94.57

Panel B: Molecular quality

Model	Clash↓ (%)	Valid yield↑ (%)	QED↑	SA↓	Diversity↑	Uniqueness↑ (%)
TSeaDiff	14.63	**66.5**	0.313 ± 0.165	5.02 ± 1.21	**0.681** $\pm$ 0.098	**99.36**
Common-only	**10.12**	58.6	**0.320** $\pm$ 0.157	4.93 ± 1.44	0.659 ± 0.135	98.23
Joint-expert	10.40	61.2	0.299 ± 0.167	**4.89** $\pm$ 1.45	0.672 ± 0.123	99.02

Molecular-glue-related properties. As shown in Panel A of Table 11, the full model achieved the highest valid yield (76.1%) and the lowest PLI profile distance (0.205), compared with 67.9% and 0.356 for the common-only variant and 71.8% and 0.236 for the joint-expert variant. It also obtained the most favorable mean full-pocket and weaker-side Vina scores, together with the highest HDR99 inclusion rate. The common-only variant showed a slightly lower clash rate (2.44% versus 3.18%), while HDR99 inclusion remained high for all three configurations. Overall, the results in Table 11 suggest that the full architecture primarily improves generation yield and agreement with the MG-PDB interaction profile, with smaller differences in HDR membership.

Molecular quality and interpretation. As shown in Panel B of Table 11, the full model achieved the highest valid yield (66.5%), diversity (0.681 ± 0.098), and uniqueness (99.36%). The common-only variant produced the lowest clash rate and the highest mean QED, whereas the joint-expert variant achieved the lowest, and therefore most favorable, SA score. These results indicate a trade-off rather than uniform improvement across all molecular-quality metrics. Nevertheless, the full architecture retains stronger generation yield and diversity while providing the clearest gains in the molecular-glue-related evaluation. Together with the architectural definitions in Appendix E.6, these ablations support the use of explicit pairwise processing and symmetric fusion, while not isolating the contribution of any individual expert or fusion component.