# TRIGLUE: A BIOLOGY-INSPIRED GENERATIVE MODEL FOR GENERATING MOLECULAR GLUE-INDUCED TERNARY COMPLEXES

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Molecular glue degraders have emerged as a promising strategy for targeted protein degradation by inducing ternary complex formation between an E3 ubiquitin ligase and a target protein. Despite their therapeutic potential, computational design of molecular glues remains largely unexplored. Unlike conventional structure-based drug design, molecular glue design is governed by the unknown protein-protein interface and requires the simultaneous modeling of ligand generation, protein-protein docking, and ternary complex assembly. In this work, we formulate molecular glue design as a ternary complex generation problem and propose a biology-inspired generative framework, **TriGlue**. Motivated by the mechanism of molecular glue action, we decompose ternary complex generation into two coupled stages: interface estimation and interface-conditioned complex generation. First, we develop an SE(3)-equivariant interface estimation module that predicts a geometrically constrained protein-protein interface from unbound monomer structures. Second, we introduce an interface-conditioned ternary flow matching network that jointly generates the molecular glue and predicts the rigid-body transformation required to assemble the ternary complex. Extensive experiments demonstrate that **TriGlue** generates chemically valid molecules and produces plausible ternary complexes, which highlight the potential of biology-inspired generative modeling for accelerating molecular glue discovery. Our code is available at `https://anonymous.4open.science/r/molecular-glue-design-806B`.

## 1 INTRODUCTION

Recently, molecular glue degraders (MGDs) have emerged as a rapidly growing strategy in targeted protein degradation (Sasso et al., 2022; Ting et al., 2024; Hsia et al., 2024; Petzold et al., 2025; Liu et al., 2025). As illustrated in Fig. 1, MGDs act through a cooperative three-body interaction among an E3 ubiquitin ligase, a molecular glue, and a target protein (POI), in which the glue mediates the E3–POI interface. The molecular glue first binds to the E3 ligase and reshapes or complements its interaction surface, creating a composite interface favorable for POI recognition. The POI is then recruited through simultaneous contacts with both the glue and the remodeled E3 surface, thereby stabilizing an otherwise weak protein–protein interaction and forming a ternary complex with a defined geometry. The resulting complex promotes POI ubiquitination and subsequent proteasomal degradation. Molecular glues are typically structurally simple, drug-like, and increasingly clinically validated (Yoon et al., 2024). Therefore, molecular glue development has become a rapidly emerging focus in modern drug discovery (Krönke et al., 2014; Lu et al., 2014; Yamamoto et al., 2020).

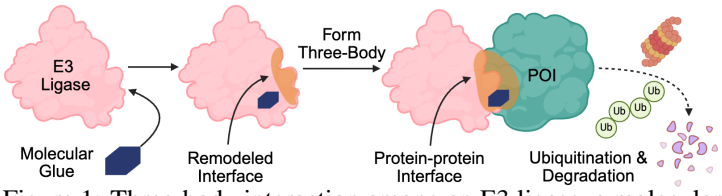



Figure 1: Three-body interaction among an E3 ligase, a molecular glue, and a POI, followed by ubiquitination and degradation.

Accelerating the discovery of molecular glues remains a central objective. Traditional molecular glue discovery has largely relied on expensive and labor-intensive experimental screening, and to date, only a few hundred molecular glues have been identified (Li et al., 2023; Huang et al.,



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.

2024). AI-based approaches have recently propelled drug discovery (Guan et al., 2023a;b; Ayadi et al., 2024; Peng et al., 2026). However, these methods are not suitable for molecular glue design, where ligand generation, protein docking, and ternary complex formulation must be considered simultaneously. Fundamentally, they are designed for binary protein–ligand binding and typically assume the existence of a predefined binding pocket on a single protein. DeepTernary (Xue et al., 2025) introduces an SE(3)-equivariant network for ternary complex structure prediction. However, it focuses on predicting structures for known ligands and does not address *de novo* drug design. Therefore, we pursue generative drug design tailored to ternary molecular glue systems.

Molecular glue design is inherently a three-body task involving an E3 ubiquitin ligase, a molecular glue, and a target protein (POI). Unlike conventional protein–ligand design, it requires a molecule that mediates interactions between two proteins and stabilizes their ternary assembly. This gives rise to two major challenges. *First*, the binding interfaces on both the E3 ligase and the POI are unknown before ternary complex formation. The model must therefore infer compatible regions on the two proteins that can support glue-mediated association, without access to a predefined binding pocket. *Second*, the molecular glue and the relative pose of the two proteins must be jointly determined. The protein arrangement shapes the space available for the glue, while the glue's geometry and chemistry determine which protein contacts can be stabilized. These interdependencies make molecular generation and protein docking tightly coupled.

To address these challenges, we propose a biology-inspired generative framework named **TriGlue**. (i) We develop an SE(3)-equivariant interface estimation module that predicts a protein–protein interface from the two monomer structures, providing a geometrically constrained and biologically meaningful structural prior. (ii) We introduce an interface-conditioned ternary flow matching network that jointly generates the molecular glue and predicts the rigid-body rotation and translation required for protein–protein docking under the guidance of the estimated interface. In summary, our main contributions are as follows:

- We investigate the novel problem of molecular glue design for ternary complexes.
- We propose a biology-inspired generative framework that decomposes ternary complex generation into interface estimation and interface-conditioned ternary flow generation.
- Extensive experiments demonstrate the effectiveness of the proposed framework in generating valid small molecules and plausible ternary complexes.

## 2 Preliminaries

### 2.1 Molecular Glue-Induced Ternary Complex Formation

As shown in Fig. 1, a molecular glue binds one protein and remodels its surface to create an interface that recruits a second partner, thereby inducing or stabilizing a protein–protein interaction. This cooperative interface combines ligand–protein and protein–protein contacts, converting transient encounters into stable ternary assemblies. Molecular glues can thereby modulate protein function, localization, or stability, which explains their growing importance in therapeutic discovery.

### 2.2 Problem Formulation

We represent a molecular-glue-induced ternary complex by its joint three-dimensional coordinates and structural attributes:

$$\mathcal{C} = ([\mathbf{X}_{\mathcal{R}}, \mathbf{X}_{\mathcal{T}}, \mathbf{X}_{\mathcal{M}}], [\mathbf{s}_{\mathcal{R}}, \mathbf{s}_{\mathcal{T}}, \mathbf{c}_{\mathcal{M}}]), \tag{1}$$

where $\mathbf{X}_{\mathcal{R}}$, $\mathbf{X}_{\mathcal{T}}$, and $\mathbf{X}_{\mathcal{M}}$ denote the coordinates of the receptor, target, and molecular glue in a shared coordinate frame. $\mathbf{s}_{\mathcal{R}}$ and $\mathbf{s}_{\mathcal{T}}$ denote the residue types of the two proteins, and $\mathbf{c}_{\mathcal{M}}$ collectively denotes the chemical attributes of the molecular glue.

**Molecular Glue-induced Protein-protein Interface.** Molecular glue activity depends on the formation of a ligand-mediated protein–protein interface (Rui et al., 2023). The ligand occupies the interfacial region and stabilizes complementary surface patches from both proteins, effectively bridging them and reducing the tendency of the complex to dissociate. Since this interface is unknown before complex assembly, we explicitly infer it as part of ternary complex generation:

$$\mathcal{I} = (\mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\mathcal{T}}), \tag{2}$$

where $\mathcal{I}_\mathcal{R}$ and $\mathcal{I}_\mathcal{T}$ characterize the interacting regions on the receptor and target, respectively. The inferred interface provides geometric guidance for both the relative assembly of the two proteins and the generation of the molecular glue.

**Dynamic Ternary Complex Assembly.** As illustrated in Fig. 2(b), ternary complex assembly couples molecular glue generation with the rigid-body motion of the target protein. Given a receptor structure $\mathcal{R}^0 = (\mathbf{s}_\mathcal{R}, \mathbf{X}^0_\mathcal{R})$ and a target structure $\mathcal{T}^0 = (\mathbf{s}_\mathcal{T}, \mathbf{X}^0_\mathcal{T})$ that are in unassembled states, our goal is to generate the corresponding ternary complex $\mathcal{C}$. Following the formulation in the preliminaries, we approximate the structural rearrangement of the two proteins during complex assembly as rigid-body transformations. Taking the receptor $\mathcal{R}^0$ as the reference frame, the transformations of the $\mathcal{R}^0$ and $\mathcal{T}^0$ in the assembling process are therefore

$$\mathbf{X}_\mathcal{R} = \mathbf{X}^0_\mathcal{R}, \qquad \mathbf{X}_\mathcal{T} = \mathbf{R}\mathbf{X}^0_\mathcal{T} + \mathbf{t}, \tag{3}$$

where $\mathbf{R} \in \mathrm{SO}(3)$ and $\mathbf{t} \in \mathbb{R}^3$ denote the rotation and translation of the target relative to the receptor. Given the above definition, the problem can be formulated as following.

**Problem 1.** *Given a receptor $\mathcal{R}^0$ and a target $\mathcal{T}^0$, we aim to generate molecular-glue-induced ternary complex by learning the target's relative transformation and the molecular glue:*

$$(\mathcal{I}, \mathbf{R}, \mathbf{t}, \mathbf{X}_\mathcal{M}, \mathbf{c}_\mathcal{M}) \sim p_\theta\left(\cdot \mid \mathcal{R}^0, \mathcal{T}^0\right), \tag{4}$$

*where $\mathcal{I} = (\mathcal{I}_\mathcal{R}, \mathcal{I}_\mathcal{T})$ denotes the molecular-glue-induced protein–protein interface, $\mathbf{R}$ and $\mathbf{t}$ denote the relative rotation and translation of $\mathcal{T}^0$ with respect to $\mathcal{R}^0$, $\mathbf{X}_\mathcal{M}$ denotes the molecular-glue coordinates in the receptor reference frame, and $\mathbf{c}_\mathcal{M}$ denotes its chemical attributes.*

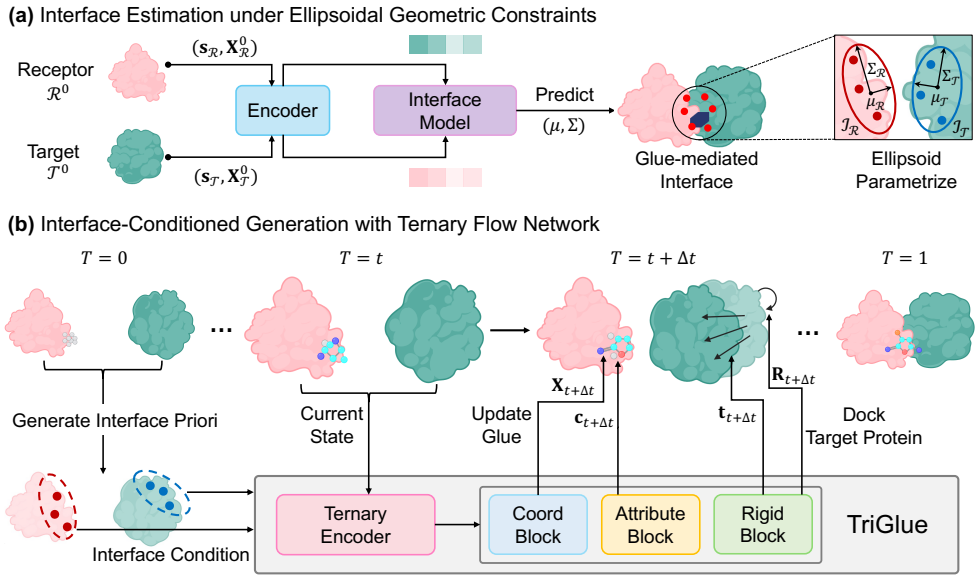



Figure 2: Overview of TriGlue. (a) The interface model predicts a glue-mediated protein–protein interface from the unassembled receptor and target, constrained by an ellipsoidal parameterization. (b) Conditioned on the predicted interface, the ternary flow network jointly updates the glue's coordinates and attributes together with the target's rigid-body pose to assemble the ternary complex.

## 3 METHOD

Following the biological process of molecular glue-induced ternary complex formation, we formulate TriGlue as a biology-inspired generative framework. As shown in Fig. 2, we factorize ternary complex generation into two coupled modules: an interface estimation module that infers a compact interface representation from two monomer proteins, and an interface-conditioned complex generation module that jointly generates the molecular glue and estimates the inter-protein rigid-body transformation to assemble a coherent ternary complex. This factorization aligns the generative process with the biological mechanism where a glue-mediated protein–protein interface guides subsequent ternary

complex assembly. The factorized design raises two key challenges. (i) the inferred interface should be constrained in a geometrically meaningful manner to ensure its utility in ternary complex modeling; (ii) the predicted interface should effectively guide a unified generative process that simultaneously produces a novel molecular glue and drives the rigid-body assembly of the two proteins. To address these challenges, we constrain the estimated interface with an ellipsoid-parameterized geometry. We further inject interface information through an interface-conditioned ternary flow network. We next describe the factorization process and the two corresponding modules in detail.

### 3.1 Biology-Inspired Factorization of Ternary Complex Generation

Inspired by the mechanism of molecular glue-induced ternary complex formation, we decompose the Problem 1 into two sub-goals: interface estimation and interface-conditioned complex generation. Specifically, this decomposition leads to two coupled probabilistic components: one estimates the molecular glue-induced interface between the two proteins, and the other generates the ternary complex conditioned on the inferred interface as follows:

$$p\big(\mathcal{I}, \mathbf{R}, \mathbf{t}, \mathbf{X}_{\mathcal{M}}, \mathbf{c}_{\mathcal{M}} \mid \mathcal{R}^0, \mathcal{T}^0\big) = \underbrace{p\big(\mathcal{I} \mid \mathcal{R}^0, \mathcal{T}^0\big)}_{\text{Interface Estimation}} \underbrace{p\big(\mathbf{R}, \mathbf{t}, \mathbf{X}_{\mathcal{M}}, \mathbf{c}_{\mathcal{M}} \mid \mathcal{I}, \mathcal{R}^0, \mathcal{T}^0\big)}_{\text{Interface-Conditioned Complex Generation}}, \tag{5}$$

where $\mathcal{I} = (\mathcal{I}_{\mathcal{R}}, \mathcal{I}_{\mathcal{T}})$ denotes the molecular glue-induced protein–protein interface. During training, $\mathcal{I}_{\mathcal{R}}$ and $\mathcal{I}_{\mathcal{T}}$ are defined as the sets of residues whose C$\alpha$ atoms lie within $8\text{Å}$ of the binding partner in the assembled complex. This biology-inspired factorization brings two key benefits: (i) It faithfully matches the mechanism of molecular-glue action; (ii) It transforms the challenging joint modeling problem into two tractable sub-tasks. We next present our approach for modeling both components.

### 3.2 Interface Estimation under Ellipsoidal Geometric Constraints

The protein–protein interface provides the geometric basis for molecular glue-mediated assembly, but is unavailable from the two unassembled structures $\mathcal{R}^0$ and $\mathcal{T}^0$. We therefore infer a compact interface representation and constrain its location and spatial extent using an ellipsoidal summary of the physical interface, as illustrated in Fig. 2(a).

**Protein Encoder**. Characterizing the interface first requires encoding each protein to obtain latent representations. We encode the two proteins $\mathcal{R}^0$ and $\mathcal{T}^0$ with an EGNN (Satorras et al., 2021):

$$(\mathbf{F}_{\mathcal{R}}, \mathbf{Z}_{\mathcal{R}}), (\mathbf{F}_{\mathcal{T}}, \mathbf{Z}_{\mathcal{T}}) = \mathrm{EGNN}(\mathcal{R}^0, \mathcal{T}^0), \tag{6}$$

where $\mathbf{F}_{\mathcal{R}} \in \mathbb{R}^{d\times N}$ and $\mathbf{F}_{\mathcal{T}} \in \mathbb{R}^{d\times M}$ are invariant residue features, while $\mathbf{Z}_{\mathcal{R}} \in \mathbb{R}^{3\times N}$ and $\mathbf{Z}_{\mathcal{T}} \in \mathbb{R}^{3\times M}$ are equivariant coordinates.

**Interface Point Representation**. Rather than assigning a hard interface label to each residue, we represent the predicted interface $\hat{\mathcal{I}}$ by two fixed-size sets of $K$ interface points (Ganea et al., 2022), $\mathbf{Y}_{\mathcal{R}}, \mathbf{Y}_{\mathcal{T}} \in \mathbb{R}^{3\times K}$. Each point is an attention-weighted aggregation of residue coordinates:

$$\mathbf{y}_{\mathcal{R},k} = \sum_{i=1}^{N} \alpha^k_{\mathcal{R},i}\, \mathbf{z}_{\mathcal{R},i}, \quad \mathbf{y}_{\mathcal{T},k} = \sum_{j=1}^{M} \alpha^k_{\mathcal{T},j}\, \mathbf{z}_{\mathcal{T},j}, \qquad k = 1, \ldots, K. \tag{7}$$

The normalized weights $\alpha^k$ are conditioned on features from both proteins, allowing different points to attend to different parts of the candidate interaction region. Together, the two point sets form a compact, differentiable representation of interface location and spatial extent. The attention parameterization is provided in Appendix D.1.

**Supervision from Interface Ellipsoids**. As illustrated in Fig. 2(a), the $K$ predicted points should collectively describe a coherent interaction region on each protein. However, these points have no canonical ordering or one-to-one correspondence with interface residues, making direct pointwise supervision ill-defined. We therefore supervise their aggregate geometry using a Gaussian ellipsoid fitted to the physical interface $\mathcal{I}$. Its centroid $\boldsymbol{\mu}$ locates the interaction region, while its covariance $\boldsymbol{\Sigma}$ captures the anisotropic extent and principal orientation. This representation permits flexible point placement while constraining the predicted interface to recover the correct region and shape. The construction of the target ellipsoid is detailed in Appendix C.

For either protein, we compute the empirical moments of the ellipsoid induced by its predicted points:

$$\hat{\boldsymbol{\mu}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{y}_k, \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{K}\sum_{k=1}^{K}(\mathbf{y}_k - \hat{\boldsymbol{\mu}})(\mathbf{y}_k - \hat{\boldsymbol{\mu}})^\top. \tag{8}$$

We train these moments to match the corresponding physical interface ellipsoid:

$$\mathcal{L}_{\text{ellip}} = \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2^2 + \|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F^2. \tag{9}$$

Applied to both proteins, the resulting ellipsoids provide compact geometric conditions for ternary generation. Their equivariance is proved in Appendix G.

### 3.3 Interface-Conditioned Generation with Ternary Flow Network

Given the predicted interface $\hat{\mathcal{I}}$, we jointly generate the molecular glue and the target pose using a ternary flow network. The generation process is conditioned on the two input proteins and their estimated interface:

$$(\mathbf{R}, \mathbf{t}, \mathbf{X}_{\mathcal{M}}, \mathbf{c}_{\mathcal{M}}) \sim p_\theta(\cdot \mid \hat{\mathcal{I}}, \mathcal{R}^0, \mathcal{T}^0). \tag{10}$$

As shown in Fig. 2(b), the network couples three flows through a shared ternary representation. (i) The *Rigid Transformation Flow* updates $(\mathbf{R}, \mathbf{t})$ to assemble the target with the fixed receptor. (ii) The *Coordinate Flow* generates ligand coordinates $\mathbf{X}_{\mathcal{M}}$. (iii) The *Attribute Flow* generates chemical attributes $\mathbf{c}_{\mathcal{M}} = (\mathbf{a}, \mathbf{b})$, comprising atom and bond types, through the Attribute block. Each flow uses the evolving state of all three components, coupling ligand generation with protein docking.

**Unified Ternary Encoder**. Modeling ternary complexes requires jointly capturing the interactions among both proteins and the ligand in a common geometric space, while preserving the proteins as rigid conditioning structures. To this end, given the input $\mathcal{G}_{\mathcal{R}}$, $\mathcal{G}_{\tilde{\mathcal{T}}}$, and $\mathcal{G}_{\tilde{\mathcal{M}}}$ (the noised graph from $\mathcal{G}_{\mathcal{M}}$), we combine all three components into a unified graph and encode them with a shared EGNN:

$$\mathbf{H}_{\mathcal{R}}, \mathbf{H}_{\tilde{\mathcal{T}}}, \mathbf{h}_{\tilde{\mathcal{M}}}, \hat{\mathbf{x}}_{\mathcal{M}} = \text{EGNN}_\theta(\mathcal{G}_{\mathcal{R}}, \mathcal{G}_{\tilde{\mathcal{T}}}, \mathcal{G}_{\tilde{\mathcal{M}}}; \Delta\mathbf{X}_{\mathcal{R}} = \Delta\mathbf{X}_{\tilde{\mathcal{T}}} = \mathbf{0}). \tag{11}$$

Here, explicit cross-entity edges allow information exchange among the receptor, target, and ligand. The encoder updates the features of all three components, yielding $\mathbf{H}_{\mathcal{R}}$, $\mathbf{H}_{\tilde{\mathcal{T}}}$, and $\mathbf{h}_{\tilde{\mathcal{M}}}$, while coordinate updates are restricted to ligand atoms to produce $\hat{\mathbf{x}}_{\mathcal{M}}$; the protein coordinates remain fixed throughout the encoding process. The denoising heads subsequently use these coupled features.

**Rigid Transformation Flow for** $p(\mathbf{R}, \mathbf{t} \mid \star)$. Predicting the target pose depends on both the interface geometry and its compatibility with the molecular glue. As illustrated in Fig. 2(b), the Rigid Transformation block predicts rotational and translational updates from the current ternary state at each timestep. These updates move the target from its pose at $t$ to that at $t + \Delta t$, progressively docking it with the fixed receptor alongside molecular glue generation.

Specifically, to construct noisy poses, we interpolate the ground-truth pose $(\mathbf{R}_1, \mathbf{t}_1)$ with a random pose, with $t = 1$ denoting the data endpoint and $t = 0$ the noise endpoint. With $\mathbf{R}_0 \sim U(\text{SO}(3))$ and $\mathbf{t}_0 \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, rotations follow a geodesic on $\text{SO}(3)$ (Chen & Lipman, 2024), while translations follow a linear path (Liu et al., 2022):

$$\mathbf{R}_t = \exp_{\mathbf{R}_0}\big(t \log_{\mathbf{R}_0}(\mathbf{R}_1)\big), \qquad \mathbf{t}_t = (1-t)\mathbf{t}_0 + t\mathbf{t}_1. \tag{12}$$

Here, $\exp(\cdot)$ and $\log(\cdot)$ are the Riemannian exponential and logarithm maps. The noisy ternary state consists of the target pose $(\mathbf{R}_t, \mathbf{t}_t)$ and ligand state $(\mathbf{X}_{\mathcal{M},t}, \mathbf{c}_{\mathcal{M},t})$. The target coordinates are $\mathbf{X}_{\mathcal{T},t} = \mathbf{R}_t\mathbf{X}_{\mathcal{T}}^0 + \mathbf{t}_t$, while the receptor remains fixed.

To denoise the target pose, the Rigid Transformation block combines the shared ternary representation with two complementary interface conditions. (i) *Coarse pose prior.* Kabsch alignment of the predicted interface points yields $(\mathbf{R}^*, \mathbf{t}^*)$ (Ganea et al., 2022), providing a reference placement that brings the two interaction regions into alignment. (ii) *Interface geometry.* At the current target pose, the relative positions of the interface points encode displacement and rotation, while their covariance matrices capture differences in the orientation and spatial extent of the two regions. This state-dependent condition quantifies the residual interface misalignment. The block integrates both interface conditions with the evolving ligand geometry to predict a vector field that guides interface alignment while accounting for protein–ligand contacts:

$$(\boldsymbol{\omega}_t, \mathbf{v}_t) = \Phi_{\text{RT}}\Big(\mathbf{R}_t, \mathbf{t}_t, \mathbf{X}_{\mathcal{M},t}, \mathbf{c}_{\mathcal{M},t}, t \mid \mathcal{R}^0, \mathcal{T}^0, \hat{\mathcal{I}}\Big), \tag{13}$$

where $\boldsymbol{\omega}_t$ is the angular field in the target's local frame and $\mathbf{v}_t$ is the translation field in the receptor reference frame. The pose prior conditions this prediction while $(\mathbf{R}_t, \mathbf{t}_t)$ remains the current flow state. The prior construction and block architecture are detailed in Appendix D.3.

For supervision, extrapolating the predicted fields over the remaining time gives endpoint estimates:

$$\widehat{\mathbf{R}}_1 = \mathbf{R}_t \exp\big((1-t)[\boldsymbol{\omega}_t\times]\big), \qquad \widehat{\mathbf{t}}_1 = \mathbf{t}_t + (1-t)\mathbf{v}_t, \tag{14}$$

Here, $[\boldsymbol{\omega}_t\times]$ is the skew-symmetric matrix representing the cross product. We train the rigid flow using a rotational vector-field loss and a translation endpoint loss:

$$\mathcal{L}_{\mathrm{RT}} = \mathbb{E}_{t\sim\mathcal{U}(0,1)}\left[\left\|\left(\log_{\mathbf{R}_t}(\widehat{\mathbf{R}}_1) - \log_{\mathbf{R}_t}(\mathbf{R}_1)\right)/(1-t)\right\|_2^2 + \|\widehat{\mathbf{t}}_1 - \mathbf{t}_1\|_2^2\right]. \tag{15}$$

At sampling time, we integrate the predicted fields in small steps alongside the coordinate and attribute flows to jointly assemble the target and molecular glue.

**Coordinate Flow**. Without a predefined pocket, the molecular glue must acquire a conformation compatible with both proteins as their relative pose evolves. We therefore denoise $\mathbf{X}_{\mathcal{M}}$ within the unified ternary EGNN, coupling ligand geometry with protein docking.

We construct noisy coordinates by linear interpolation, $\mathbf{X}_{\mathcal{M},t} = (1-t)\mathbf{X}_{\mathcal{M},0} + t\mathbf{X}_{\mathcal{M},1}$, where each atom in $\mathbf{X}_{\mathcal{M},0}$ is sampled from $\mathcal{N}(\mathbf{0}, \mathbf{I}_3)$ and $\mathbf{X}_{\mathcal{M},1}$ contains the ground-truth coordinates. To denoise this state, the Coordinate block aggregates interactions among ligand atoms and with both proteins at the current target pose $(\mathbf{R}_t, \mathbf{t}_t)$. As the coordinate branch of the shared encoder, it updates only ligand coordinates and predicts the endpoint:

$$\widehat{\mathbf{X}}_{\mathcal{M},1} = \Phi_{\mathbf{X}}\big(\mathbf{R}_t, \mathbf{t}_t, \mathbf{X}_{\mathcal{M},t}, \mathbf{c}_{\mathcal{M},t}, t \mid \mathcal{R}^0, \mathcal{T}^0\big). \tag{16}$$

Both proteins remain fixed within each encoder evaluation, with the target pose updated separately by the rigid flow. The coordinate objective is

$$\mathcal{L}_{\mathrm{coor}} = \mathbb{E}_{t\sim\mathcal{U}(0,1)}\left[\|\widehat{\mathbf{X}}_{\mathcal{M},1} - \mathbf{X}_{\mathcal{M},1}\|_F^2\right]. \tag{17}$$

The corresponding sampling field is $(\widehat{\mathbf{X}}_{\mathcal{M},1} - \mathbf{X}_{\mathcal{M},t})/(1-t)$.

**Attribute Flow**. The glue's atom and bond types must be compatible with its evolving conformation and protein environment. We therefore jointly denoise the chemical attributes $\mathbf{c}_{\mathcal{M}} = (\mathbf{a}, \mathbf{b})$ using the shared ternary representation. To construct continuous noisy states, we encode both atom and bond types as scaled soft one-hot logits (Li et al., 2024), collectively denoted by $\mathbf{c}_{\mathcal{M},1}$. Their interpolation path is $\mathbf{c}_{\mathcal{M},t} = (1-t)\mathbf{c}_{\mathcal{M},0} + t\mathbf{c}_{\mathcal{M},1}$, with $\mathbf{c}_{\mathcal{M},0} \sim \mathcal{N}(\mathbf{0}, \sigma_c^2\mathbf{I})$ and noise scale $\sigma_c$. The Attribute block uses the shared features to predict atom types and combines atom-pair features with the denoised ligand geometry to predict undirected bond types as follows:

$$\widehat{\mathbf{c}}_{\mathcal{M},1} = \Phi_{\mathrm{attr}}\big(\mathbf{R}_t, \mathbf{t}_t, \mathbf{X}_{\mathcal{M},t}, \mathbf{c}_{\mathcal{M},t}, t \mid \mathcal{R}^0, \mathcal{T}^0\big). \tag{18}$$

We supervise both attribute components through

$$\mathcal{L}_{\mathrm{attr}} = \mathbb{E}_{t\sim\mathcal{U}(0,1)}\left[\|\widehat{\mathbf{c}}_{\mathcal{M},1} - \mathbf{c}_{\mathcal{M},1}\|_2^2\right], \tag{19}$$

where the squared norm sums the atom- and bond-logit errors. At sampling time, the field $(\widehat{\mathbf{c}}_{\mathcal{M},1} - \mathbf{c}_{\mathcal{M},t})/(1-t)$ updates the logits, which are decoded into atom and bond types at the endpoint. The two prediction branches are detailed in Appendix D.5.

**Training Objective**. Finally, we balance the gradients of all flows using weighting coefficients $\lambda_i$ and jointly train the entire ternary denoising framework as follows:

$$\mathcal{L}_{\text{flow}} = \lambda_1\mathcal{L}_{\mathrm{RT}} + \lambda_2\mathcal{L}_{\mathrm{coor}} + \lambda_3\mathcal{L}_{\mathrm{attr}}. \tag{20}$$

We provide the details of the model architecture and the flow matching in the Appendix D and E.

**Sampling**. We provide the detailed sampling algorithm in the Appendix F.

# 4 EXPERIMENT

In this section, we conduct experiments to answer the following research questions:

- **RQ1:** Can TriGlue generate high-quality molecular glues while accurately docking the two proteins into a reasonable ternary complex?
- **RQ2:** How accurate and geometrically meaningful are the ellipsoid interfaces predicted by TriGlue?
- **RQ3:** How does interface conditioning affect the performance of TriGlue?

## 4.1 EXPERIMENTAL SETUP

**Dataset**. We utilize **TernaryDB** (Xue et al., 2025) for training and evaluation, which contains 22,303 ternary complexes curated from the Protein Data Bank (Berman et al., 2000), covering diverse proteins and drug-like ligands. To prevent data leakage, complexes are clustered by protein sequence similarity using MMseqs2 (Steinegger & Söding, 2017), and clusters containing known molecular glue complexes are excluded from training, ensuring disjoint validation and test splits.

**Baselines**. Since no method jointly performs *de novo* molecular glue generation and ternary complex reconstruction, we evaluate the two sub-tasks separately. Molecular generation baselines include **AR** (Luo et al., 2021), **TargetDiff** (Guan et al., 2023a), and **PocketXMol** (Peng et al., 2026); protein-protein docking baselines include **DeepTernary** (Xue et al., 2025), **EquiDock** (Ganea et al., 2022), and **DiffDock-PP** (Ketata et al., 2023).

**Training Details**. For the interface modeling module, the EGNN encoder consists of 4 layers with a hidden dimension of 128. The model is trained using the AdamW optimizer (Loshchilov & Hutter, 2019) with an initial learning rate of $5 \times 10^{-4}$ and a cosine annealing schedule for 1200 epochs. For the ternary flow network, the unified EGNN consists of 4 layers with a hidden dimension of 128. We train it for 800 epochs using Adam with a linear learning rate decay from $5 \times 10^{-4}$ to $3 \times 10^{-4}$. The weights for translation, rotation, atom-type, bond-type, coordinate, and pose-coordinate losses are $\{1, 1, 1, 1, 5, 1\}$, respectively. All experiments are conducted on 8 NVIDIA L40 GPUs.

**Evaluation Metrics**. We evaluate performance from two perspectives. For molecular generation, we report AutoDock Vina-based affinity metrics (**Vina**, **Vina min**, **Vina dock**) (Trott & Olson, 2010), the high-affinity ratio (**Aff**), and molecular property metrics (**QED**, **SA**). For ternary-complex reconstruction, we measure ligand pose accuracy using **RMSD** and protein docking quality using **DockQ** (Basu & Wallner, 2016) and its success rate (DockQ $> 0.23$).

Dataset statistics, baseline descriptions, training details, and metric definitions are provided in Appendices H.1, H.2, H.3, and H.4, respectively.

Table 1: Performance comparison on generated molecules.

| Method | RMSD (↓) | Vina score (↓) | | Vina min (↓) | | Vina dock (↓) | | QED (↑) | | SA (↑) | | Aff (↑) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Avg | Avg | Med | Avg | Med | Avg | Med | Avg | Med | Avg | Med | Avg |
| AR | $8.16_{2.17}$ | 8.76 | 9.04 | 3.62 | 3.69 | -4.39 | -4.87 | 0.51 | 0.54 | 0.54 | 0.53 | 0.53 |
| TargetDiff | $6.27_{1.99}$ | -2.11 | -6.09 | -4.83 | -6.74 | -7.55 | -7.86 | 0.47 | 0.53 | 0.55 | 0.58 | 0.65 |
| PocketXMol | $6.74_{0.91}$ | -6.17 | -6.14 | -6.88 | -6.88 | -7.69 | -7.65 | 0.47 | 0.49 | 0.79 | 0.78 | 0.69 |
| TriGlue | $4.95_{1.15}$ | -6.57 | -7.12 | -7.19 | -7.76 | -8.68 | -9.38 | 0.72 | 0.79 | 0.74 | 0.78 | 0.82 |

## 4.2 TERNARY COMPLEX GENERATION

Since no existing baseline jointly performs molecular-glue generation and ternary-complex reconstruction end-to-end, we answer **RQ1** from two complementary perspectives: the quality of the generated molecules and the accuracy of protein docking. Under this evaluation framework, we first assess zero-shot generalization to unseen molecular-glue families and then explore whether a few family-specific examples can further improve the pretrained model.

**Zero-shot generation.** (i) For molecular generation, despite receiving no binding-pocket information, TriGlue achieves the strongest or tied-strongest result in 11 of the 12 columns in Tab. 1, including the lowest average RMSD (4.95), the strongest Vina results, and the highest QED and high-affinity

ratio. PocketXMol remains strongest in average SA and ties TriGlue in median SA. (ii) For protein docking, TriGlue substantially outperforms the baselines in Fig. 3(a,b), achieving a median DockQ of 0.22 and an acceptable-docking rate of 47.3%, versus 0.09 and 31.5% for the strongest baseline, DeepTernary. **Few-shot extension.** Starting from the zero-shot checkpoint, we fine-tune TriGlue on a few support complexes per family while holding out all query complexes. As shown in Fig. 3(c), adaptation improves the mean DockQ in 10 of 13 families by 0.05 on average, with the largest gains on TIR1 (+0.22), CaM (+0.16), and $\beta$-TrCP (+0.12). Although gains vary across families with limited support data, these results show that few-shot supervision can effectively refine the transferable geometric prior for specific families. Additional results are provided in Appendix I.2.

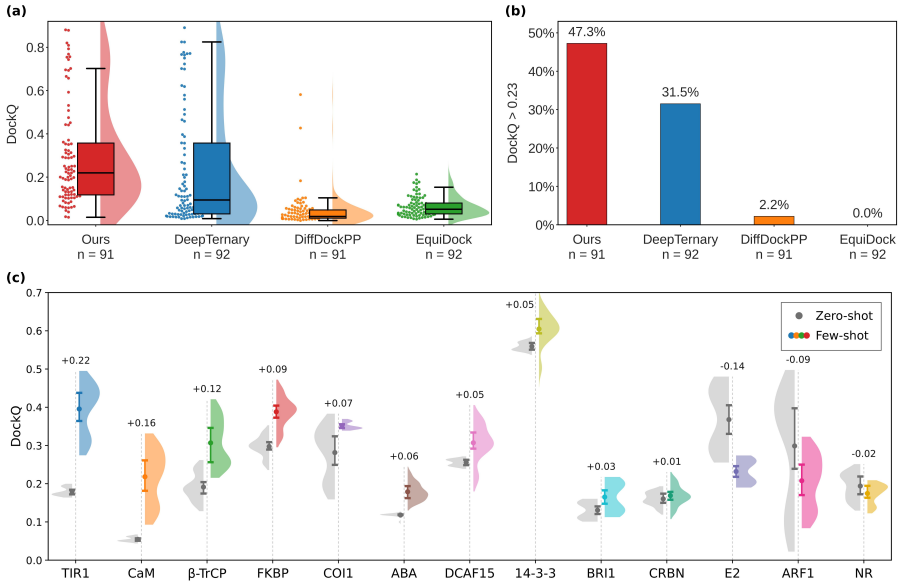



Figure 3: Zero- and few-shot protein docking results. (a) DockQ distributions and (b) acceptable-docking rates (DockQ $>$ 0.23) under the zero-shot setting. (c) Family-wise DockQ distributions before (gray) and after (colored) few-shot adaptation; annotations denote changes in mean DockQ.

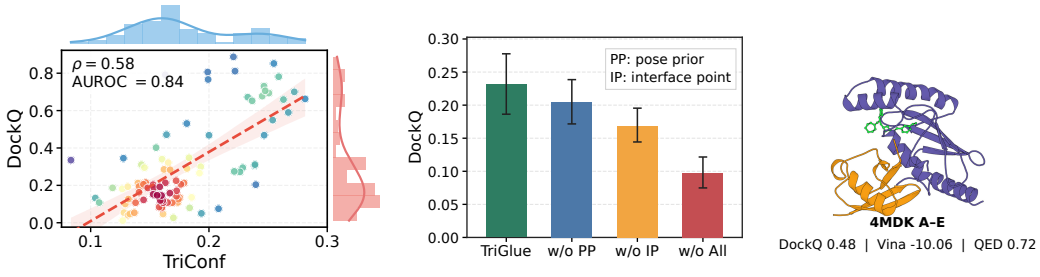
**Robustness to predicted apo structures.** To evaluate a practical setting where an experimentally resolved receptor is unavailable, we replace it with a receptor-only AlphaFold3 prediction (Abramson et al., 2024), while retaining the experimentally determined target structure. As detailed in Appendix Sec. I.3 and Tab. 5, across 92 paired test complexes, the mean best DockQ remains nearly unchanged (0.309 vs. 0.305), while the mean best ligand RMSD increases by only 0.08 Å. These results suggest that TriGlue is robust to realistic structural uncertainty in apo receptor inputs.

### 4.3 Ellipsoid-Parameterized Interface Prediction

To further answer **RQ2**, we evaluate estimated interface ellipsoids on the held-out TernaryDB test split and compare against **PeSTo** (Krapp et al., 2023) and **ScanNet** (Tubiana et al., 2022). Residue-level interface predictions from these methods are converted into ellipsoids. We report geometry- and distribution-aware metrics, including JSD, Wasserstein-2 distance (W2), Center-L2, and LogDet error (see Appendix H.5). As shown in Tab. 2, our method consistently outperforms both baselines across all metrics, achieving substantially lower divergence and improved geometric alignment with ground-truth interface ellipsoids.

Table 2: Ellipsoid interface estimation on the test set.

| Method | JSD ↓ | W2 ↓ | Center-L2 ↓ | LogDet ↓ |
|---|---|---|---|---|
| Ours | 0.47 ± 0.15 | 12.65 ± 7.89 | 10.99 ± 7.47 | 3.83 ± 5.01 |
| PeSTo | 0.68 ± 0.04 | 58.26 ± 46.03 | 57.39 ± 46.25 | 8.00 ± 11.54 |
| ScanNet | 0.68 ± 0.04 | 58.96 ± 45.34 | 57.57 ± 45.85 | 10.20 ± 12.92 |




(a) Indicator score *TriConf* (b) Interface-condition ablations (c) A sampled case

Figure 4: Interface-derived confidence, conditioning ablations, and a representative TriGlue design.

### 4.4 TriConf: Emergent Confidence from Interface Model

Beyond its role in conditioning generation, the interface predictor yields an unexpected byproduct: a confidence signal for ternary-complex quality. TriConf measures how consistently the two interface ellipsoids are predicted under small perturbations of the input proteins and maps their median $W_2$ deviation to $[0, 1]$, where a higher value indicates greater interface stability (Appendix H.6). It is computed using a single batched interface-model evaluation and requires neither a ground-truth complex nor full generative sampling. As shown in Fig. 4a, across 91 test complexes, TriConf correlates positively with DockQ (Spearman $\rho = 0.58$, $p = 1.9 \times 10^{-9}$) and achieves an AUROC of 0.84 for identifying acceptable docking, indicating that TriConf produced by the interface model provides a GT-free approximation of ternary-complex quality.

**Ablation Study**. To answer **RQ3**, we ablate the two interface-conditioning routes in $\Phi_{\mathrm{RT}}$: the Kabsch pose prior and the interface-point geometry. As shown in Fig. 4b, TriGlue achieves a mean best-of-20 DockQ of $0.231$, compared with $0.204$, $0.169$, and $0.097$ after removing the pose prior, interface points, and all interface conditions, respectively. The larger degradation caused by removing the interface points, and especially both routes, demonstrates that explicit interface geometry is important for recovering the ternary pose. Detailed experimental settings are provided in Appendix I.1.

## 5 Related Work

**Geometric generative modeling.** Flow matching learns continuous transport from a simple prior to a data distribution (Lipman et al., 2023) and extends naturally to geometric domains through Riemannian formulations (Chen & Lipman, 2024). Together with equivariant diffusion, it has enabled protein backbone, binder, and peptide generation (Bose et al., 2024; Yim et al., 2023; Watson et al., 2023; Ahern et al., 2025; Stark et al., 2025; Lin et al., 2024; Li et al., 2024). Structure predictors such as AlphaFold-Multimer and AlphaFold3 model multi-chain assemblies (Evans et al., 2021; Abramson et al., 2024), but operate on specified molecular components rather than discovering both a small-molecule mediator and the geometry of a new three-body interaction.

**Molecular and complex design.** Structure-based models generate ligands within known binding pockets (Guan et al., 2023a; Peng et al., 2026; Zhang et al., 2026), while protein docking methods predict pairwise poses (Ganea et al., 2022; Ketata et al., 2023) and DeepTernary predicts ternary structures for specified components (Xue et al., 2025). In contrast, TriGlue addresses an unknown protein–protein interface and jointly generates the molecular glue and relative protein pose required to assemble the ternary complex. An extended discussion is provided in Appendix B.

## 6 Conclusion

Molecular glue design is fundamentally a three-body generation problem: ligand chemistry, the protein–protein interface, and the relative protein pose must be resolved together. We introduced TriGlue, which predicts an SE(3)-equivariant geometric interface from unbound proteins and uses it to jointly generate a molecular glue and assemble the ternary complex. Experiments show strong molecular generation and docking performance, transferable zero-shot behavior, and targeted gains from few-shot adaptation. By turning the unknown interface from a hidden constraint into an explicit generative condition, TriGlue offers a unified route toward computational molecular glue discovery.

### AI USE STATEMENT

Generative AI tools were used only for grammar checking and limited language polishing. They were not used to develop the research ideas, methods, code, experiments, data analysis, or scientific claims. The authors reviewed all suggested edits and take full responsibility for the final content of this paper.

### ETHICS STATEMENT

This study uses publicly available molecular structures and does not involve human participants or personal data. TriGlue produces computational candidates for molecular glue research. The generated compounds and predicted ternary complexes have not been experimentally or clinically validated. Any downstream synthesis or biological testing should follow appropriate laboratory safety and institutional review procedures.

### REPRODUCIBILITY STATEMENT

We provide the source code in an anonymized repository at `https://anonymous.4open.science/r/molecular-glue-design-806B`.

## REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

Woody Ahern, Jason Yim, Doug Tischer, Saman Salike, Seth M Woodbury, Donghyo Kim, Indrek Kalvet, Yakov Kipnis, Brian Coventry, Han Raut Altae-Tran, et al. Atom-level enzyme active site scaffolding using rfdiffusion2. *Nature Methods*, pp. 1–10, 2025.

Sirine Ayadi, Leon Hetzel, Johanna Sommer, Fabian J Theis, and Stephan Günnemann. Unified guidance for geometry-conditioned molecular generation. In *Neural Information Processing Systems (NeurIPS)*, 2024.

Steven M Banik, Kayvon Pedram, Simon Wisnovsky, Green Ahn, Nicholas M Riley, and Carolyn R Bertozzi. Lysosome-targeting chimaeras for degradation of extracellular proteins. *Nature*, 2020.

Sankar Basu and Björn Wallner. Dockq: a quality measure for protein-protein docking models. *PloS one*, 2016.

Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic Acids Research*, 28(1): 235–242, 2000.

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, et al. $\pi_0$: A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024.

Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian FATRAS, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael M. Bronstein, and Alexander Tong. SE(3)-stochastic flow matching for protein backbone generation. In *International Conference on Learning Representations (ICLR)*, 2024.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations (ICLR)*, 2024.

Adam D Cotton, Duy P Nguyen, Josef A Gramespacher, Ian B Seiple, and James A Wells. Development of antibody-based protacs for the degradation of the cell-surface immune checkpoint protein pd-l1. *Journal of the American Chemical Society*, 2021.

Robin CE Deutscher, Christian Meyners, Maximilian L Repity, Wisely Oki Sugiarto, Jürgen M Kolos, Edvaldo VS Maciel, Tim Heymann, Thomas M Geiger, Stefan Knapp, Frederik Lermyte, et al. Discovery of fully synthetic fkbp12-mtor molecular glues. *Chemical Science*, 2025.

Jeffrey A Dewey, Clémence Delalande, Saara-Anne Azizi, Vivian Lu, Dionysios Antonopoulos, and Gyorgy Babnigg. Molecular glue discovery: current and future approaches. *Journal of Medicinal Chemistry*, 2023.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *International Conference on Machine Learning (ICML)*, 2024.

Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, et al. Protein complex prediction with alphafold-multimer. *biorxiv*, 2021.

Octavian-Eugen Ganea, Xinyuan Huang, Charlotte Bunne, Yatao Bian, Regina Barzilay, Tommi S Jaakkola, and Andreas Krause. Independent se (3)-equivariant models for end-to-end rigid protein docking. In *International Conference on Learning Representations (ICLR)*, 2022.

Shanshan Gu, Danrui Cui, Xiaoyu Chen, Xiufang Xiong, and Yongchao Zhao. Protacs: an emerging targeting technique for protein degradation in drug discovery. *BioEssays*, 2018.

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

Jiaqi Guan, Xiangxin Zhou, Yuwei Yang, Yu Bao, Jian Peng, Jianzhu Ma, Qiang Liu, Liang Wang, and Quanquan Gu. Decompdiff: Diffusion models with decomposed priors for structure-based drug design. In *International Conference on Machine Learning (ICML)*. PMLR, 2023b.

Oliver Hsia, Matthias Hinterndorfer, Angus D Cowan, Kentaro Iso, Tasuku Ishida, Ramasubramanian Sundaramoorthy, Mark A Nakasone, Hana Imrichova, Caroline Schätz, Andrea Rukavina, et al. Targeted protein degradation via intramolecular bivalent glues. *Nature*, 627(8002):204–211, 2024.

Hai-Tsang Huang, Ryan J Lumpkin, Ryan W Tsai, Shuyao Su, Xu Zhao, Yuan Xiong, James Chen, Nada Mageed, Katherine A Donovan, Eric S Fischer, et al. Ubiquitin-specific proximity labeling for the identification of e3 ligase substrates. *Nature Chemical Biology*, 20(9):1227–1236, 2024.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

Mohamed Amine Ketata, Cedrik Laue, Ruslan Mammadov, Hannes Stark, Menghua Wu, Gabriele Corso, Céline Marquet, Regina Barzilay, and Tommi S Jaakkola. Diffdock-pp: Rigid protein-protein docking with diffusion models. In *ICLR 2023-Machine Learning for Drug Discovery Workshop*, 2023.

Xiangzhe Kong, Rui Jiao, Haowei Lin, Ruihan Guo, Wenbing Huang, Wei-Ying Ma, Zihua Wang, Yang Liu, and Jianzhu Ma. Peptide design through binding interface mimicry with pepmimic. *Nature Biomedical Engineering*, pp. 1–16, 2025.

Lucien F Krapp, Luciano A Abriata, Fabio Cortés Rodriguez, and Matteo Dal Peraro. Pesto: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. *Nature Communications*, 14(1):2175, 2023.

Jan Krönke, Namrata D Udeshi, Anupama Narla, Peter Grauman, Slater N Hurst, Marie McConkey, Tanya Svinkina, Dirk Heckl, Eamon Comer, Xiaoyu Li, et al. Lenalidomide causes selective degradation of ikzf1 and ikzf3 in multiple myeloma cells. *Science*, 2014.

Alice Shi Ming Li, Serah Kimani, Brian Wilson, Mahmoud Noureldin, Héctor González-Álvarez, Ahmed Mamai, Laurent Hoffer, John P Guilinger, Ying Zhang, Moritz von Rechenberg, et al. Discovery of nanomolar dcaf1 small molecule ligands. *Journal of Medicinal Chemistry*, 2023.

Jiahan Li, Chaoran Cheng, Zuofan Wu, Ruihan Guo, Shitong Luo, Zhizhou Ren, Jian Peng, and Jianzhu Ma. Full-atom peptide design based on multi-modal flow matching. In *International Conference on Machine Learning (ICML)*, 2024.

Haitao Lin, Odin Zhang, Huifeng Zhao, Dejun Jiang, Lirong Wu, Zicheng Liu, Yufei Huang, and Stan Z Li. Ppflow: Target-aware peptide design with torsional flow matching. In *International Conference on Machine Learning (ICML)*, 2024.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

Xingchao Liu, Chengyue Gong, et al. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2022.

Yanfen Liu, Jieyun Bai, Dong Li, and Yong Cang. Routes to molecular glue degrader discovery. *Trends in Biochemical Sciences*, 2025.

AEA Lopez-Girona, D Mendy, T Ito, K Miller, AK Gandhi, J Kang, S Karasawa, G Carmel, P Jackson, M Abbasian, et al. Cereblon is a direct protein target for immunomodulatory and antiproliferative activities of lenalidomide and pomalidomide. *Leukemia*, 2012.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Gang Lu, Richard E Middleton, Huahang Sun, MarkVic Naniong, Christopher J Ott, Constantine S Mitsiades, Kwok-Kin Wong, James E Bradner, and William G Kaelin Jr. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of ikaros proteins. *Science*, 2014.

Shitong Luo, Jiaqi Guan, Jianzhu Ma, and Jian Peng. A 3d generative model for structure-based drug design. In *Neural Information Processing Systems (NeurIPS)*, 2021.

Hossam Nada, Yongseok Choi, Sungdo Kim, Kwon Su Jeong, Nicholas A Meanwell, and Kyeong Lee. New insights into protein–protein interaction modulators in drug discovery and therapeutic advance. *Signal Transduction and Targeted Therapy*, 2024.

Xingang Peng, Ruihan Guo, Fenglin Guo, Ziyi Wang, Jiayu Sun, Jiaqi Guan, Yinjun Jia, Yan Xu, Yanwen Huang, Muhan Zhang, et al. Unified modeling of 3d molecular generation via atomic interactions with pocketxmol. *Cell*, 2026.

Georg Petzold, Pablo Gainza, Stefano Annunziato, Ilaria Lamberto, Peter Trenh, Laura A McAllister, Bradley DeMarco, Laura Schwander, Richard D Bunker, Mary Zlotosch, et al. Mining the crbn target space redefines rules for molecular glue–induced neosubstrate recognition. *Science*, 389 (6755):eadt6736, 2025.

Adam Polyak, Amit Zohar, Andrew Brown, Andros Tjandra, Animesh Sinha, Ann Lee, Apoorv Vyas, Bowen Shi, Chih-Yao Ma, Ching-Yao Chuang, et al. Movie gen: A cast of media foundation models. *arXiv preprint arXiv:2410.13720*, 2024.

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.

Kusal TG Samarasinghe and Craig M Crews. Targeted protein degradation: A promise for undruggable proteins. *Cell Chemical Biology*, 2021.

Janet M Sasso, Rumiana Tenchov, DaSheng Wang, Linda S Johnson, Xinmei Wang, and Qiongqiong Angela Zhou. Molecular glues: the adhesive connecting targeted protein degradation to the clinic. *Biochemistry*, 2022.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International Conference on Machine Learning (ICML)*, 2021.

Matthieu Schapira, Matthew F Calabrese, Alex N Bullock, and Craig M Crews. Targeted protein degradation: expanding the toolbox. *Nature Reviews Drug Discovery*, 2019.

Stuart L Schreiber. The rise of molecular glues. *Cell*, 2021.

Hannes Stark, Bowen Jing, Chenyu Wang, Gabriele Corso, Bonnie Berger, Regina Barzilay, and Tommi Jaakkola. Dirichlet flow matching with applications to dna sequence design. In *International Conference on Machine Learning (ICML)*, pp. 46495–46513, 2024.

Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy John O'Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, Mateo Reveiz, Roman Bushuiev, Tomáš Pluskal, Josef Sivic, Karsten Kreis, Arash Vahdat, Shamayeeta Ray, Jonathan T. Goldstein, Andrew Savinov, Jacob A. Hambalek, Anshika Gupta, Diego A. Taquiri-Diaz, Yaotian Zhang, A. Katherine Hatstat, Angelika Arada, Nam Hyeong Kim, Ethel Tackie-Yarboi, Dylan Boselli, Lee Schnaider, Chang C. Liu, Gene-Wei Li, Denes Hnisz, David M. Sabatini, William F. DeGrado, Jeremy Wohlwend, Gabriele Corso, Regina Barzilay, and Tommi Jaakkola. Boltzgen: Toward universal binder design. *bioRxiv*, 2025.

Martin Steinegger and Johannes Söding. Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028, 2017.

Pamela Y Ting, Sneha Borikar, John Ryan Kerrigan, Noel M Thomsen, Eamon Aghania, Amelia E Hinman, Alejandro Reyes, Nicolas Pizzato, Barna D Fodor, Fabian Wu, et al. A molecular glue degrader of the wiz transcription factor for fetal hemoglobin induction. *Science*, 2024.

O Trott and AJ 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.

Jérôme Tubiana, Dina Schneidman-Duhovny, and Haim J Wolfson. Scannet: an interpretable geometric deep learning model for structure-based protein binding site prediction. *Nature Methods*, 19(6):730–739, 2022.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research (JMLR)*, 9(86):2579–2605, 2008.

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

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.

Junichi Yamamoto, Tetsufumi Suwa, Yuki Murase, Shumpei Tateno, Hirotaka Mizutome, Tomoko Asatsuma-Okumura, Nobuyuki Shimizu, Tsutomu Kishi, Shuji Momose, Masahiro Kizaki, et al. Arid2 is a pomalidomide-dependent crl4crbn substrate in multiple myeloma cells. *Nature Chemical Biology*, 2020.

Haijuan Yang, Derek G Rudge, Joseph D Koos, Bhamini Vaidialingam, Hyo J Yang, and Nikola P Pavletich. mtor kinase structure, mechanism and regulation. *Nature*, 2013.

Jason Yim, Brian L Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. Se (3) diffusion model with application to protein backbone generation. In *International Conference on Machine Learning (ICML)*, 2023.

Hojong Yoon, Justine C Rutter, Yen-Der Li, Benjamin L Ebert, et al. Induced protein degradation for therapeutics: past, present, and future. *The Journal of Clinical Investigation*, 2024.

Xudong Zhang, Sanqing Qu, Fan Lu, Jianmin Wang, Zhixin Tian, Shangding Gu, Yanping Zhang, Alois Knoll, Shaorong Gao, Guang Chen, and Changjun Jiang. Steering semi-flexible molecular diffusion model for structure-based drug design with reinforcement learning. *Science Advances*, 12(16), 2026. doi: 10.1126/sciadv.ady9955. URL https://www.science.org/doi/10.1126/sciadv.ady9955.

Lin Zhao, Jia Zhao, Kunhong Zhong, Aiping Tong, and Da Jia. Targeted protein degradation: mechanisms, strategies and application. *Signal Transduction and Targeted Therapy*, 2022.

Guangcai Zhong, Xiaoyu Chang, Weilin Xie, and Xiangxiang Zhou. Targeted protein degradation: advances in drug discovery and clinical practice. *Signal Transduction and Targeted Therapy*, 2024.

# APPENDIX CONTENTS

**A Terminology** **15**

**B Extended Related Work** **16**

**C Ellipsoid Construction** **17**

**D Model Architecture** **17**

D.1 Interface Point Attention . . . . . 17

D.2 Unified Ternary EGNN . . . . . 17

D.3 Rigid Transformation Block . . . . . 18

D.4 Coordinate Block . . . . . 21

D.5 Attribute Block . . . . . 21

**E Flow Matching** **21**

**F Sample Algorithm** **23**

**G Proofs** **24**

**H Experimental Details** **25**

H.1 Dataset Details . . . . . 25

H.2 Baseline Details . . . . . 25

H.3 Training Details . . . . . 26

H.4 Evaluation Metric Details . . . . . 26

H.5 Ellipsoid Interface Prediction Metric Details . . . . . 28

H.6 TriConf Definition . . . . . 29

**I Additional Experiment Results** **29**

I.1 Interface Ablation Details . . . . . 29

I.2 Ternary Interaction Recovery after Family Adaptation . . . . . 30

I.3 Robustness to Predicted Apo Structures . . . . . 30

I.4 Interface Representation Analysis . . . . . 31

**J Family-Wise Case Studies of Ternary-Complex Design** **31**

# A TERMINOLOGY

Table 3 provides supplementary introductions to the biological background and abbreviations of the technical terms that appear in this paper.

Table 3: Terminology list.

| Terminology | Abbreviation | Description |
|---|---|---|
| **Protein of Interest** | **POI** | The specific protein intended to be modulated, inhibited, stabilized, or degraded by a therapeutic molecule or biological mechanism in order to achieve a desired biological or therapeutic effect. |
| **Targeted Protein Degradation** | **TPD** | A therapeutic strategy that eliminates disease-related proteins by harnessing the cell's natural protein disposal machinery rather than inhibiting protein activity directly. |
| **Proteasome** | – | A large multisubunit protease complex that degrades ubiquitin-tagged proteins and maintains protein quality control in cells. |
| **Molecular Glue Degrader** | **MGD** | A small molecule that induces interaction between an E3 ligase and a target protein, promoting ubiquitination and subsequent proteasomal degradation. |
| **Proteolysis Targeting Chimeras** | **PROTACs** | Heterobifunctional molecules that bind a protein of interest via a target warhead and recruit an E3 ligase via a second ligand, connected by a chemical linker that controls ternary complex geometry. |
| **E3 Ligase** | – | An enzyme that recognizes substrate proteins and catalyzes ubiquitin transfer from E2 enzymes, determining substrate specificity in the ubiquitin proteasome system. |
| **Ubiquitin** | – | A conserved regulatory protein that can be attached to substrates to mark them for degradation or regulate signaling pathways. |
| **Ubiquitination** | – | A post-translational modification in which ubiquitin, a small regulatory protein, is covalently attached to lysine residues on a substrate protein through an enzymatic cascade involving E1, E2, and E3 enzymes, commonly marking the protein for proteasomal degradation or altering its cellular function. |

Traditional drug discovery has long relied on inhibitors to modulate protein activity directly, yet this approach faces limitations when targeting certain disease-related proteins. Targeted protein degradation (TPD) (Gu et al., 2018; Schreiber, 2021; Banik et al., 2020; Cotton et al., 2021) has emerged as a transformative therapeutic paradigm, which harnesses the cell's intrinsic proteolytic machinery, such as the proteasome and lysosomal pathways, to selectively degrade disease-associated proteins, in contrast to traditional small-molecule inhibitors that merely block protein function (Schapira et al., 2019; Samarasinghe & Crews, 2021; Zhao et al., 2022; Zhong et al., 2024).

Here, we provide additional information related to molecular glues.

**Molecular Glue** is a class of small molecules that modulate protein–protein interactions by enhancing the affinity between specific proteins within cells. These compounds can induce novel protein interactions or stabilize existing ones, providing an alternative therapeutic approach to traditional drug

discovery. Notably, molecular glues offer a means to target proteins once considered "undruggable" by conventional strategies. They function through diverse mechanisms, e.g., promoting targeted protein degradation or disrupting essential protein functions, and hold potential for treating a range of diseases, including cancer and neurodegenerative disorders.

Molecular glues can be broadly divided according to their functional outcome (Schreiber, 2021; Dewey et al., 2023):

- **Degradative molecular glues** recruit a protein of interest to an E3 ubiquitin ligase, promoting its ubiquitination and subsequent proteasomal degradation. Lenalidomide, for example, binds cereblon and induces the degradation of neo-substrates such as IKZF1 and IKZF3 (Zhao et al., 2022).
- **Non-degradative molecular glues** induce or stabilize protein–protein interactions without targeting a component for degradation. Depending on the recruited partners, the resulting complex can inhibit, activate, or otherwise alter protein function.

This work focuses specifically on degradative molecular glues that induce E3 ligase–POI ternary complexes.

Generally, Molecular Glues employ two primary mechanisms to modulate protein-protein interactions: allosteric regulation and direct bridging.

- **Allosteric Regulation**. In the context of allosteric regulation, molecular glues interact with a primary protein, triggering conformational rearrangements that generate or reinforce novel binding interfaces. These newly formed interfaces facilitate the recruitment of a secondary protein (Nada et al., 2024). E.g., lenalidomide associates with the E3 ubiquitin ligase cereblon (CRBN), altering its structural configuration to enable the recruitment and subsequent ubiquitination of neo-substrates like IKZF1 and IKZF3, leading to their degradation (Lopez-Girona et al., 2012).
- **Direct Bridging**. Unlike allosteric modulation, direct bridging entails the molecular glue simultaneously binding to two proteins at their interface, effectively connecting them. A classic example is rapamycin, which binds both FKBP12 and mTOR to form a stable ternary complex, thereby suppressing mTOR's kinase function (Yang et al., 2013). Although such direct bridging mechanisms do occur, they are relatively rare compared to allosteric modulation. The latter is more prevalent among molecular glues because it leverages the inherent flexibility of protein conformations to induce new interaction surfaces, even in the absence of pre-defined binding sites (Deutscher et al., 2025).

## B Extended Related Work

**Flow Matching and Geometric Generation**. Flow matching learns a continuous-time vector field that transports samples from a simple prior to the data distribution (Lipman et al., 2023). It has shown strong performance in Euclidean domains, including image (Esser et al., 2024), video (Polyak et al., 2024), and robotic action generation (Black et al., 2024). Extensions to non-Euclidean spaces provide a principled way to model structured data on Riemannian manifolds (Chen & Lipman, 2024), including protein backbones (Bose et al., 2024) and biological sequences (Stark et al., 2024). Related equivariant diffusion models directly respect three-dimensional symmetry and have been widely used for protein backbone and complex generation (Yim et al., 2023; Watson et al., 2023; Ahern et al., 2025). Our work combines Euclidean, simplex, and Riemannian flows to model ligand coordinates, atom types, and protein rigid transformations within one ternary generation process.

**Protein Structure and Complex Design**. The success of AlphaFold (Jumper et al., 2021) stimulated rapid progress in learning protein structures and assemblies. AlphaFold-Multimer (Evans et al., 2021) and AlphaFold3 (Abramson et al., 2024) extend structure prediction to multi-chain and biomolecular complexes, while generative methods address protein backbones, enzymes, binders, and peptides (Watson et al., 2023; Ahern et al., 2025; Yim et al., 2023; Stark et al., 2025; Lin et al., 2024; Li et al., 2024; Kong et al., 2025). These approaches primarily generate or predict protein structures for specified molecular components. By comparison, molecular glue design must discover a small molecule while simultaneously determining how two initially unbound proteins assemble around it.

**Structure-Based Molecular Generation and Docking**. Structure-based molecular design methods such as AR (Luo et al., 2021), TargetDiff (Guan et al., 2023a), and PocketXMol (Peng et al., 2026) generate ligands conditioned on a known protein pocket. Protein–protein docking methods, including EquiDock (Ganea et al., 2022) and DiffDock-PP (Ketata et al., 2023), instead predict the relative pose of two proteins but do not design the molecule mediating their interaction. DeepTernary (Xue et al., 2025) directly predicts ternary complex structures from specified components, yet it does not perform *de novo* molecular glue generation. TriGlue bridges these directions by first estimating the unknown protein–protein interface and then jointly generating the glue and rigid-body transformation, treating molecular glue discovery as a coupled three-body design problem.

# C Ellipsoid Construction

Given an interface residue set $\mathcal{I}$ (either $\mathcal{I}_{\mathcal{R}}$ or $\mathcal{I}_{\tilde{\mathcal{T}}}$), we approximate its spatial distribution using a Gaussian ellipsoid parameterized by a centroid $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. This representation provides a compact description of the interface location, spatial extent, and principal orientation in 3D space. Formally, the centroid is computed as the mean of the C$\alpha$ coordinates:

$$\boldsymbol{\mu} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \mathbf{X}_i, \tag{21}$$

and the covariance matrix is estimated as:

$$\boldsymbol{\Sigma} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} (\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^\top, \tag{22}$$

where $\mathbf{X}_i \in \mathbb{R}^3$ denotes the C$\alpha$ coordinate of residue $i$.

The resulting Gaussian ellipsoid $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ captures the global geometry of the interface: the mean encodes the interface center, while the eigenstructure of $\boldsymbol{\Sigma}$ reflects its spatial spread and principal axes. This compact parameterization serves as the geometric representation used throughout our framework for interface alignment and evaluation.

# D Model Architecture

This section provides architectural details for the interface estimator and the ternary flow network. We first specify the attention used to construct the interface points. We then describe the shared equivariant encoder and the rigid transformation block used during ternary generation.

## D.1 Interface Point Attention

For each interface point $k$, the attention weights over one protein are conditioned on a pooled representation of the other protein:

$$\begin{aligned} \alpha^k_{\mathcal{R},i} &= \mathrm{softmax}_i \left( \frac{\mathbf{f}^\top_{\mathcal{R},i} \mathbf{W}^k_{\mathcal{R}} \phi(\mathbf{F}_{\mathcal{T}})}{\sqrt{d}} \right), \\ \alpha^k_{\mathcal{T},j} &= \mathrm{softmax}_j \left( \frac{\mathbf{f}^\top_{\mathcal{T},j} \mathbf{W}^k_{\mathcal{T}} \phi(\mathbf{F}_{\mathcal{R}})}{\sqrt{d}} \right), \end{aligned} \tag{23}$$

where $\mathbf{W}^k_{\mathcal{R}}, \mathbf{W}^k_{\mathcal{T}} \in \mathbb{R}^{d \times d}$ are learned projections and $\phi(\cdot)$ denotes average pooling. The construction is symmetric across the two proteins.

## D.2 Unified Ternary EGNN

**Common coordinate frame.** The receptor remains in the reference frame, whereas the unbound target is placed according to the current flow state:

$$\mathbf{X}_{\mathcal{T},t} = \mathbf{X}_{\tilde{\mathcal{T}}} \mathbf{R}_t^\top + \mathbf{1} \mathbf{t}_t^\top. \tag{24}$$

The initial coordinate array is the concatenation

$$\mathbf{X}^{(0)} = \mathrm{Concat}[\mathbf{X}_{\mathcal{R}}, \mathbf{X}_{\mathcal{T},t}, \mathbf{X}_{\mathcal{M},t}]\,. \tag{25}$$

Only C$\alpha$ coordinates are used for protein nodes in the shared graph. Before message passing, all valid coordinates are centered by their masked global mean $\mathbf{c}$ and scaled by $s_x$:

$$\widetilde{\mathbf{X}}^{(0)} = (\mathbf{X}^{(0)} - \mathbf{1}\mathbf{c}^{\top})/s_x. \tag{26}$$

This normalization is compatible with global rigid transformations and stabilizes distance features when the current target translation is large.

**Initial node and edge features.** For each protein residue, the input feature contains a learned amino-acid embedding, masked heavy-atom coordinates expressed in its N–C$\alpha$–C local frame, and angular encodings of the three backbone dihedrals. For each ligand atom, the current noisy atom type $\mathbf{a}_{i,t}$ is mapped through a learned embedding and a two-layer MLP. The three entity-specific feature streams are projected to a common hidden dimension and augmented with a shared sinusoidal timestep embedding and a learned entity label:

$$\mathbf{h}_i^{(0)} = \mathrm{LN}\big(\mathbf{W}_{e(i)}\mathbf{s}_i + \mathbf{W}_t\mathbf{e}(t) + \mathbf{e}_{e(i)}\big)\,, \qquad e(i) \in \{\mathcal{R}, \mathcal{T}, \mathcal{M}\}. \tag{27}$$

The noisy bond type $\mathbf{b}_{ij,t}$ is used as a categorical edge token on ligand–ligand pairs. Protein–protein and cross-entity pairs receive the zero edge token, corresponding to no chemical bond. Thus, the evolving bond state affects both the ligand messages and the final bond prediction.

**Cross-entity neighborhoods.** A purely global $k$-nearest-neighbor graph can omit an entire component when the three entities are far apart early in the flow. We therefore explicitly reserve $k_c$ directed neighbors from each of the other two entity types for every valid receiver node. The remaining neighborhood slots are selected by geometric proximity over the unified graph. Consequently, message passing contains directed receptor–target, receptor–ligand, and target–ligand edges in both directions, in addition to within-entity geometric neighbors.

**Equivariant message passing.** Let $\mathcal{N}(i)$ denote the resulting neighborhood and let $\boldsymbol{\epsilon}_{ij}$ be the embedding of the edge token. At layer $\ell$, messages and node features are updated as

$$\mathbf{m}_{ij}^{(\ell)} = \phi_e^{(\ell)}\Big(\mathbf{h}_i^{(\ell)}, \mathbf{h}_j^{(\ell)}, \|\widetilde{\mathbf{x}}_i^{(\ell)} - \widetilde{\mathbf{x}}_j^{(\ell)}\|_2^2, \boldsymbol{\epsilon}_{ij}\Big)\,, \tag{28}$$

$$\overline{\mathbf{m}}_i^{(\ell)} = \mathrm{Mean}_{j\in\mathcal{N}(i)}\,\mathbf{m}_{ij}^{(\ell)}, \tag{29}$$

$$\mathbf{h}_i^{(\ell+1)} = \mathbf{h}_i^{(\ell)} + \phi_h^{(\ell)}\Big(\mathbf{h}_i^{(\ell)}, \overline{\mathbf{m}}_i^{(\ell)}\Big)\,. \tag{30}$$

Each message also predicts an invariant scalar coordinate weight $w_{ij}^{(\ell)}$. Coordinates are updated along normalized relative directions:

$$\widetilde{\mathbf{x}}_i^{(\ell+1)} = \widetilde{\mathbf{x}}_i^{(\ell)} + \chi_i \sum_{j\in\mathcal{N}(i)} w_{ij}^{(\ell)}\,\mathrm{Norm}\Big(\widetilde{\mathbf{x}}_i^{(\ell)} - \widetilde{\mathbf{x}}_j^{(\ell)}\Big)\,, \qquad \chi_i = \begin{cases} 1, & e(i) = \mathcal{M}, \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

The mask $\chi_i$ permits coordinate updates only for valid ligand atoms. Receptor and target nodes exchange messages with the entire graph and update their features, but their coordinates are returned exactly unchanged. After $L$ layers, the network produces the ligand coordinate prediction $\widehat{\mathbf{x}}_{\mathcal{M},1}$ and the coupled representations $\mathbf{H}_{\mathcal{R}}$, $\mathbf{H}_{\tilde{\mathcal{T}}}$, and $\mathbf{h}_{\tilde{\mathcal{M}}}$. We use hidden dimension 128, $L = 4$, total neighborhood size 32, $k_c = 8$, message dimension 64, and coordinate scale $s_x = 10$ Å.

### D.3 Rigid Transformation Block

The Rigid Transformation block $\Phi_{\mathrm{RT}}$ parameterizes a ligand-conditioned vector field on $\mathrm{SO}(3) \times \mathbb{R}^3$. Its base point is always the current noisy pose $(\mathbf{R}_t, \mathbf{t}_t)$; the Kabsch solution $(\mathbf{R}^*, \mathbf{t}^*)$ is used only to construct geometric conditions and never replaces the current state. The block consumes the coupled node representations and coordinates produced by the unified ternary EGNN, together with the virtual interfaces, their covariance matrices, and the timestep embedding. It returns a normalized world-frame translation field $\widetilde{\mathbf{v}}_t$ and a right-trivialized, body-frame angular field $\boldsymbol{\omega}_t$.

**Kabsch pose prior.** We align the predicted target interface points to the receptor interface points to obtain a coarse relative-pose condition. Using the row-coordinate convention in this appendix, let $\mathbf{Y}_{\mathcal{R}}, \mathbf{Y}_{\tilde{\mathcal{T}}} \in \mathbb{R}^{K\times 3}$ contain the points before target assembly, and let $\bar{\mathbf{Y}}_{\mathcal{R}}, \bar{\mathbf{Y}}_{\tilde{\mathcal{T}}}$ denote their centered versions. The Kabsch solution is

$$\begin{aligned} \bar{\mathbf{Y}}_{\tilde{\mathcal{T}}}^{\top}\bar{\mathbf{Y}}_{\mathcal{R}} = \mathbf{U}\mathbf{S}\mathbf{V}^{\top}, \quad \mathbf{D} = \operatorname{diag}\big(1, 1, \det(\mathbf{V}\mathbf{U}^{\top})\big), \\ \mathbf{R}^{*} = \mathbf{V}\mathbf{D}\mathbf{U}^{\top}, \quad \mathbf{t}^{*} = \boldsymbol{\mu}_{\mathcal{R}}^{Y} - \mathbf{R}^{*}\boldsymbol{\mu}_{\tilde{\mathcal{T}}}^{Y}, \end{aligned} \tag{32}$$

where the centroids $\boldsymbol{\mu}_{\mathcal{R}}^{Y}, \boldsymbol{\mu}_{\tilde{\mathcal{T}}}^{Y} \in \mathbb{R}^3$ are column vectors. Thus the aligned target points are $\mathbf{Y}_{\tilde{\mathcal{T}}}\mathbf{R}^{*\top} + \mathbf{1}\mathbf{t}^{*\top}$.

**Current ternary geometry and pooled contexts.** Let $\mathbf{X}_{\mathcal{R}} = \{\mathbf{x}_{\mathcal{R},i}\}$ denote receptor coordinates, $\mathbf{X}_{\mathcal{T},t} = \{\mathbf{x}_{\mathcal{T},i,t}\}$ the current target coordinates, and $\widehat{\mathbf{X}}_{\mathcal{M},1} = \{\widehat{\mathbf{x}}_{\mathcal{M},a,1}\}$ the ligand coordinates predicted by the shared EGNN. The target virtual interface follows the exact rigid-flow state,

$$\mathbf{Y}_{\mathcal{T},t} = \mathbf{Y}_{\tilde{\mathcal{T}}}\mathbf{R}_t^{\top} + \mathbf{1}\mathbf{t}_t^{\top}, \tag{33}$$

whereas the same $\mathbf{X}_{\mathcal{T},t}$ is passed through the shared EGNN without coordinate refinement: target features are updated by ternary message passing, while the coordinate-update mask keeps all protein nodes fixed. For a masked node set $\mathcal{V}$, define $\operatorname{Pool}(\mathbf{H}) = |\mathcal{V}|^{-1}\sum_{i\in\mathcal{V}}\mathbf{h}_i$. The invariant ternary and time contexts are

$$\mathbf{c}_{\mathrm{tri}} = \phi_{\mathrm{tri}}(\operatorname{Concat}\left[\operatorname{Pool}(\mathbf{H}_{\mathcal{R}}), \operatorname{Pool}(\mathbf{H}_{\tilde{\mathcal{T}}}), \operatorname{Pool}(\mathbf{h}_{\tilde{\mathcal{M}}})\right]), \tag{34}$$

$$\mathbf{c}_t = \phi_t(\operatorname{Pool}(\mathbf{e}_t)), \tag{35}$$

where $\mathbf{e}_t$ is the sinusoidal timestep embedding and both $\phi_{\mathrm{tri}}$ and $\phi_t$ are two-layer MLPs.

**Learned protein–ligand contact summaries.** To expose the global head to ligand-conditioned geometry, we construct one normalized contact distribution for receptor–ligand pairs and one for target–ligand pairs. For $P \in \{\mathcal{R}, \mathcal{T}\}$,

$$\ell_{ia}^{P} = \frac{\langle \mathbf{W}_P\mathbf{h}_{P,i}, \mathbf{W}_M\mathbf{h}_{\mathcal{M},a}\rangle}{\sqrt{d_c}} + \phi_d^{P}(\mathrm{RBF}(\|\mathbf{x}_{P,i} - \widehat{\mathbf{x}}_{\mathcal{M},a,1}\|_2)), \tag{36}$$

$$q_{ia}^{P} = \frac{m_i^{P} m_a^{\mathcal{M}} \exp(\ell_{ia}^{P})}{\sum_{j,b} m_j^{P} m_b^{\mathcal{M}} \exp(\ell_{jb}^{P})}. \tag{37}$$

The corresponding soft contact points are

$$\mathbf{p}_P = \sum_{i,a} q_{ia}^{P}\mathbf{x}_{P,i}, \qquad \mathbf{p}_{M\leftarrow P} = \sum_{i,a} q_{ia}^{P}\widehat{\mathbf{x}}_{\mathcal{M},a,1}. \tag{38}$$

Thus, $\mathbf{p}_{M\leftarrow\mathcal{R}}$ and $\mathbf{p}_{M\leftarrow\mathcal{T}}$ can differ because they summarize the ligand atoms most relevant to each protein.

**Global equivariant bases.** Let $\boldsymbol{\mu}_{\mathcal{R}}^{Y}$ and $\boldsymbol{\mu}_{\mathcal{T},t}^{Y}$ be the virtual-interface centroids; $\mathbf{c}_{\mathcal{R}}$ and $\mathbf{c}_{\mathcal{T},t}$ the receptor and current-target centroids; $\widehat{\mathbf{c}}_{\mathcal{M},1}$ the predicted ligand centroid; and $\mathbf{c}_{\mathcal{T}}^{*} = \mathbf{c}_{\tilde{\mathcal{T}}}\mathbf{R}^{*\top} + \mathbf{t}^{*}$ the target centroid under the Kabsch prior. We first define the following interface-, contact-, and pose-dependent geometric vectors:

$$\mathbf{d}_Y = \frac{1}{K}\sum_k(\mathbf{y}_{\mathcal{R},k} - \mathbf{y}_{\mathcal{T},k,t}), \qquad \boldsymbol{\tau}_Y = \frac{1}{K}\sum_k(\mathbf{y}_{\mathcal{T},k,t} - \boldsymbol{\mu}_{\mathcal{T},t}^{Y}) \times (\mathbf{y}_{\mathcal{R},k} - \mathbf{y}_{\mathcal{T},k,t}), \tag{39}$$

$$\mathbf{d}_M = \mathbf{p}_{M\leftarrow\mathcal{T}} - \mathbf{p}_{\mathcal{T}}, \qquad \boldsymbol{\tau}_M = \sum_{i,a} q_{ia}^{\mathcal{T}}(\mathbf{x}_{\mathcal{T},i,t} - \mathbf{c}_{\mathcal{T},t}) \times (\widehat{\mathbf{x}}_{\mathcal{M},a,1} - \mathbf{x}_{\mathcal{T},i,t}), \tag{40}$$

$$\mathbf{d}_N = \mathbf{p}_{M\leftarrow\mathcal{R}} - \mathbf{p}_{\mathcal{T}}, \qquad \mathbf{d}_C = \widehat{\mathbf{c}}_{\mathcal{M},1} - \mathbf{c}_{\mathcal{T},t}, \tag{41}$$

$$\mathbf{d}_K = \mathbf{c}_{\mathcal{T}}^{*} - \mathbf{c}_{\mathcal{T},t}, \qquad \boldsymbol{\rho}_K = \mathbf{R}_t\big(\log(\mathbf{R}_t^{\top}\mathbf{R}^{*})\big)^{\vee}, \tag{42}$$

$$\mathbf{u}_{\mathcal{R}} = \boldsymbol{\mu}_{\mathcal{R}}^{Y} - \mathbf{c}_{\mathcal{R}}, \qquad \mathbf{u}_{\mathcal{T}} = \boldsymbol{\mu}_{\mathcal{T},t}^{Y} - \mathbf{c}_{\mathcal{T},t}. \tag{43}$$

Here, $(\cdot)^\vee$ maps a skew-symmetric matrix to its rotation vector. From these quantities, we construct six world-frame translation bases and six world-frame rotation bases:

$$\mathcal{B}_{\mathrm{trans}} = [\mathbf{d}_Y, \mathbf{d}_M, \mathbf{d}_N, \mathbf{d}_C, \mathbf{d}_K, \mathbf{p}_{\mathcal{R}} - \mathbf{p}_{\mathcal{T}}], \tag{44}$$

$$\mathcal{B}_{\mathrm{rot}} = [\boldsymbol{\tau}_Y, \boldsymbol{\tau}_M, (\mathbf{p}_{\mathcal{T}} - \mathbf{c}_{\mathcal{T},t}) \times \mathbf{d}_M, \boldsymbol{\rho}_K, \mathbf{u}_{\mathcal{T}} \times \mathbf{u}_{\mathcal{R}}, (\mathbf{p}_{\mathcal{T}} - \mathbf{c}_{\mathcal{T},t}) \times \mathbf{d}_N]. \tag{45}$$

The translation bases respectively encode virtual-interface alignment, target–ligand contact, receptor-conditioned ligand contact, ligand-centroid displacement, the Kabsch center prior, and receptor–target contact alignment. The rotation bases encode the corresponding interface, ligand-contact, Kabsch, and cross-contact torques. The network predicts one invariant coefficient for each basis. In particular, let $\mathbf{z}_t \in \mathbb{R}^{12}$ concatenate

$$\begin{aligned}&\big[\|\mathbf{d}_Y\|^2, \|\mathbf{d}_M\|^2, \|\mathbf{d}_N\|^2, \|\mathbf{d}_C\|^2, \|\mathbf{d}_K\|^2, \|\big(\log(\mathbf{R}_t^\top \mathbf{R}^*)\big)^\vee\|^2,\\ &\mathbb{E}_{q^{\mathcal{R}}} d_{ia}, \mathbb{E}_{q^{\mathcal{T}}} d_{ia}, \|\boldsymbol{\Sigma}_{\mathcal{R}} - \mathbf{R}_t \boldsymbol{\Sigma}_{\tilde{\mathcal{T}}} \mathbf{R}_t^\top\|_F^2, \mathbf{d}_Y^\top \mathbf{d}_M, \mathbf{d}_M^\top \mathbf{d}_K, \mathbf{u}_{\mathcal{R}}^\top \mathbf{u}_{\mathcal{T}}\big].\end{aligned} \tag{46}$$

After applying the elementwise signed logarithm $\mathrm{slog}(z) = \mathrm{sign}(z)\log(1+|z|)$, an MLP produces twelve invariant weights:

$$[\mathbf{a}^{\mathrm{trans}}, \mathbf{a}^{\mathrm{rot}}] = \phi_g(\mathrm{Concat}[\mathrm{slog}(\mathbf{z}_t), \mathbf{c}_{\mathrm{tri}}, \mathbf{c}_t]), \qquad \mathbf{a}^{\mathrm{trans}}, \mathbf{a}^{\mathrm{rot}} \in \mathbb{R}^6. \tag{47}$$

**Local force–torque field.** The global bases are complemented by residue/atom-level interactions on directed edges from each target residue $i$ to its $k$ nearest receptor residues and $k$ nearest ligand atoms. For an edge $(i,j)$ whose endpoint type is $e(j) \in \{\mathcal{R}, \mathcal{M}\}$, the shared edge MLP predicts $H$ pairs of invariant coefficients:

$$[\alpha_{ij}^{1:H}, \beta_{ij}^{1:H}] = \phi_e\big(\mathrm{Concat}[\mathbf{W}_{\mathcal{T}}\mathbf{h}_{\mathcal{T},i}, \mathbf{W}_{e(j)}\mathbf{h}_j, \mathrm{RBF}(d_{ij}), \mathbf{c}_t, \mathbf{e}_{e(j)}]\big), \tag{48}$$

$$\widehat{\mathbf{r}}_{ij} = \frac{\mathbf{x}_j - \mathbf{x}_{\mathcal{T},i,t}}{d_{ij}}, \qquad \overline{\mathbf{r}}_i = \frac{\mathbf{x}_{\mathcal{T},i,t} - \mathbf{c}_{\mathcal{T},t}}{r_{\mathcal{T}}}, \tag{49}$$

$$\mathbf{f}_{ij}^h = \alpha_{ij}^h \widehat{\mathbf{r}}_{ij}, \qquad \boldsymbol{\tau}_{ij}^h = \beta_{ij}^h(\overline{\mathbf{r}}_i \times \widehat{\mathbf{r}}_{ij}), \tag{50}$$

where $r_{\mathcal{T}}$ is the masked root-mean-square target radius. For a receptor endpoint, $\mathbf{x}_j = \mathbf{x}_{\mathcal{R},j}$; for a ligand endpoint, $\mathbf{x}_j = \widehat{\mathbf{x}}_{\mathcal{M},j,1}$. Each head is mean-aggregated over valid edges, and the heads and endpoint types are combined as

$$\mathbf{f}_{\mathrm{loc}} = \frac{1}{\sqrt{H}} \sum_{h=1}^{H} (\overline{\mathbf{f}}_{\mathcal{R}}^h + \overline{\mathbf{f}}_{\mathcal{M}}^h), \qquad \boldsymbol{\tau}_{\mathrm{loc}} = \frac{1}{\sqrt{H}} \sum_{h=1}^{H} (\overline{\boldsymbol{\tau}}_{\mathcal{R}}^h + \overline{\boldsymbol{\tau}}_{\mathcal{M}}^h). \tag{51}$$

**Vector-field output and pose update.** Let $\mathbf{b}_q^{\mathrm{trans}}$ and $\mathbf{b}_q^{\mathrm{rot}}$, $q = 1, \ldots, 6$, denote the $q$-th entries of $\mathcal{B}_{\mathrm{trans}}$ and $\mathcal{B}_{\mathrm{rot}}$, respectively. Writing $\widehat{\mathbf{b}} = \mathbf{b}/\max(\|\mathbf{b}\|_2, \epsilon)$ for safe radial normalization, the network weights these global bases with $\mathbf{a}^{\mathrm{trans}}$ and $\mathbf{a}^{\mathrm{rot}}$ and combines them with the local force–torque field:

$$\widetilde{\mathbf{v}}_t = \sum_{q=1}^{6} a_q^{\mathrm{trans}} \widehat{\mathbf{b}}_q^{\mathrm{trans}} + \mathbf{f}_{\mathrm{loc}}, \tag{52}$$

$$\boldsymbol{\omega}_t^{\mathrm{world}} = \sum_{q=1}^{6} a_q^{\mathrm{rot}} \widehat{\mathbf{b}}_q^{\mathrm{rot}} + \boldsymbol{\tau}_{\mathrm{loc}}, \tag{53}$$

$$\boldsymbol{\omega}_t = \mathbf{R}_t^\top \boldsymbol{\omega}_t^{\mathrm{world}}, \qquad \mathbf{v}_t = s_{\mathrm{trans}} \widetilde{\mathbf{v}}_t. \tag{54}$$

The implementation uses $H = 8$, $k = 8$, and the fixed conversion scale $s_{\mathrm{trans}} = 50$ Å; it radially clips $\|\boldsymbol{\omega}_t\|_2$ to at most $\pi$. The translation field is learned in normalized units and converted to physical units only by $s_{\mathrm{trans}}$. The endpoint estimates used during training are

$$\widehat{\mathbf{R}}_1 = \mathbf{R}_t \exp((1-t)[\boldsymbol{\omega}_t\times]), \qquad \widehat{\mathbf{t}}_1 = \mathbf{t}_t + (1-t)\mathbf{v}_t. \tag{55}$$

At sampling step $n$ with step size $\Delta t_n$, the same vector field is integrated without resetting to the Kabsch prior:

$$\mathbf{R}_{n+1} = \mathrm{Proj}_{\mathrm{SO}(3)}(\mathbf{R}_n \exp(\Delta t_n[\boldsymbol{\omega}_n\times])), \qquad \mathbf{t}_{n+1} = \mathbf{t}_n + \Delta t_n \mathbf{v}_n. \tag{56}$$

Because all learned coefficients are invariant scalars and all vector bases transform equivariantly, $\mathbf{v}_t$ and $\boldsymbol{\omega}_t^{\mathrm{world}}$ rotate with a global coordinate transformation, while the body-frame $\boldsymbol{\omega}_t$ remains consistent with the right-multiplicative rotation update.

### D.4 COORDINATE BLOCK

The Coordinate block $\Phi_{\mathbf{X}}$ is the coordinate-update branch of the unified ternary EGNN. It progressively refines the noisy ligand coordinates within every message-passing layer using both ligand–ligand and protein–ligand edges. The mask in Eq. 31 keeps the two proteins fixed, so the ligand state is the only learned geometric output. Mapping the final state back to the physical coordinate frame gives

$$\widehat{\mathbf{x}}_{i,1} = s_x \widetilde{\mathbf{x}}_i^{(L)} + \mathbf{c}, \qquad i \in \mathcal{V}_{\mathcal{M}}. \tag{57}$$

Because the scalar coordinate weights are invariant and the relative directions are equivariant, $\Phi_{\mathbf{X}}$ is SE(3)-equivariant. Its prediction $\widehat{\mathbf{X}}_{\mathcal{M},1}$ is subsequently used by both the Attribute block and $\Phi_{\mathrm{RT}}$, coupling ligand geometry to bond inference and rigid-pose prediction.

### D.5 ATTRIBUTE BLOCK

The Attribute block $\Phi_{\mathrm{attr}}$ contains an atom branch and a bond branch. The noisy atom state $\mathbf{a}_t$ enters the initial ligand node features, while the noisy bond state $\mathbf{b}_t$ enters the ligand–ligand edge features. Consequently, both categorical states participate in all layers of ternary message passing before their final predictions are read out.

**Atom-type prediction.** For each valid ligand atom, the atom branch applies layer normalization followed by a two-layer MLP to its final representation:

$$\hat{\mathbf{s}}_{i,1}^{\mathrm{a}} = \mathbf{W}_2^{\mathrm{a}} \,\mathrm{SiLU}\Big(\mathbf{W}_1^{\mathrm{a}} \,\mathrm{LN}(\mathbf{h}_i^{(L)})\Big). \tag{58}$$

The output contains one logit for each of the 25 atom classes; predictions at padded positions are masked.

**Bond-type prediction.** For every ligand atom pair $(i, j)$, the bond branch combines the final node representations, the geometry predicted by $\Phi_{\mathbf{X}}$, and the current noisy bond state. We use the symmetric pair representation

$$\mathbf{z}_{ij} = \mathrm{Concat}\left[\mathbf{h}_i^{(L)} + \mathbf{h}_j^{(L)}, |\mathbf{h}_i^{(L)} - \mathbf{h}_j^{(L)}|, \frac{\|\widehat{\mathbf{x}}_{i,1} - \widehat{\mathbf{x}}_{j,1}\|_2^2}{s_x^2}, \mathbf{E}_b(\mathbf{b}_{ij,t})\right], \qquad \hat{\mathbf{s}}_{ij,1}^{\mathrm{b}} = \phi_{\mathrm{b}}(\mathbf{z}_{ij}). \tag{59}$$

Here, $\mathbf{E}_b$ is the bond embedding shared with the EGNN edge tokens, and $\phi_{\mathrm{b}}$ is a layer-normalized two-layer SiLU MLP. Its six logits correspond to no bond, single, double, triple, aromatic, and other bonds. We additionally enforce undirected-bond symmetry by

$$\hat{\mathbf{s}}_{ij,1}^{\mathrm{b}} \leftarrow \frac{1}{2}\left(\hat{\mathbf{s}}_{ij,1}^{\mathrm{b}} + \hat{\mathbf{s}}_{ji,1}^{\mathrm{b}}\right), \tag{60}$$

Self-pairs and padded pairs are excluded from the prediction. The two branches together constitute $\Phi_{\mathrm{attr}}$ and are jointly supervised by the Attribute Flow objective in Sec. E.

## E FLOW MATCHING

Throughout this appendix, we abbreviate the generation conditions as $\star := (\hat{\mathcal{I}}, \mathcal{R}^0, \mathcal{T}^0, \mathbf{R}^*, \mathbf{t}^*)$. Because $(\mathbf{R}^*, \mathbf{t}^*)$ is deterministically obtained from $\hat{\mathcal{I}}$ by Kabsch alignment, this is equivalent to conditioning on $(\hat{\mathcal{I}}, \mathcal{R}^0, \mathcal{T}^0)$ in the main text.

**Conditional Flow Matching**. The conditional flow matching (CFM) framework (Lipman et al., 2023) offers an approach to learn a probability time-dependent flow $\psi_t : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d$ that transforms samples from a simple base distribution $p_0$ into a target data distribution $p_1 \approx q$. The flow and its corresponding vector field $u_t : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d$ are governed by the ordinary differential equation (ODE): $\frac{\mathrm{d}}{\mathrm{d}t} x_t = u_t(x_t)$, where $x_t = \psi_t(x_0)$. However, it is challenging to model this vector field directly due to the complexity of its closed-form solution. Instead, CFM circumvents this limitation by conditioning the flow and vector field on a specific data pair: a sample $x_0 \sim p_0$ and a target

$x_1 \sim p_1$. This leads to a tractable formulation of the interpolated trajectory $x_t = \psi_t(x_0 \mid x_1)$ and the conditional vector field $u_t(x_t \mid x_1) = \frac{\mathrm{d}}{\mathrm{d}t}x_t$. Hence, the standard CFM training loss is formulated as:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t\sim\mathcal{U}(0,1)} \left\| v_t(\psi_t(x_0 \mid x_1)) - \frac{\mathrm{d}}{\mathrm{d}t}x_t \right\|_2^2, \tag{61}$$

where $\mathcal{U}(0,1)$ is a uniform distribution and $v_t$ is a neural network parameterized by $\theta$.

Based on the CFM, our Ternary Flow Network involves three types of flow:

**Euclidean CFM for Coordinate.** The Euclidean conditional flow interpolates between Gaussian noise $\mathbf{x}_{i,0} \sim \mathcal{N}(\mathbf{0}, I_3)$ and the data distribution $\mathbf{x}_{i,1} \sim p(\mathbf{x}_i \mid \star)$:

$$\psi_t(\mathbf{x}_{i,0} \mid \mathbf{x}_{i,1}) = t\mathbf{x}_{i,1} + (1-t)\mathbf{x}_{i,0} \tag{62}$$

The target vector field is:

$$u_t = \frac{d}{dt}\psi_t = \mathbf{x}_{i,1} - \mathbf{x}_{i,0}. \tag{63}$$

The Coordinate block predicts the clean endpoint $\widehat{\mathbf{x}}_{i,1}$, which induces the sampling field $(\widehat{\mathbf{x}}_{i,1} - \mathbf{x}_{i,t})/(1-t)$. Its endpoint objective is given below.

**Simplex Paths for Attributes**. The same construction is used for the two categorical variables in the Attribute Flow. Let $\mathbf{u}$ denote either an atom type $\mathbf{a}_i$ or an undirected bond type $\mathbf{b}_{ij}$, with $d_{\mathbf{u}}$ classes. Following PepFlow (Li et al., 2024), its label is mapped to a scaled soft one-hot endpoint $\mathbf{s}_1^{\mathbf{u}} \in \mathbb{R}^{d_{\mathbf{u}}}$:

$$\mathbf{s}_1^{\mathbf{u}}[k] = \begin{cases} K, & k = \mathbf{u}, \\ -K, & \text{otherwise.} \end{cases} \tag{64}$$

Collecting these logits over all atoms and unordered bond pairs gives the continuous attribute representation $\mathbf{c}_{\mathcal{M},1}$ used in the main text, with $K = \sigma_c$. Its softmax for each variable parameterizes a point on $\Delta^{d_{\mathbf{u}}-1}$. We define the conditional flow in logit space between $\mathbf{s}_0^{\mathbf{u}} \sim \mathcal{N}(\mathbf{0}, K^2\mathbf{I})$ and $\mathbf{s}_1^{\mathbf{u}}$ as

$$\boldsymbol{\psi}_t(\mathbf{s}_0^{\mathbf{u}} \mid \mathbf{s}_1^{\mathbf{u}}) = t\mathbf{s}_1^{\mathbf{u}} + (1-t)\mathbf{s}_0^{\mathbf{u}}, \qquad \dot{\mathbf{s}}_t^{\mathbf{u}} = \mathbf{s}_1^{\mathbf{u}} - \mathbf{s}_0^{\mathbf{u}}. \tag{65}$$

The Attribute block predicts the endpoint class distributions from $\mathbf{s}_t^{\mathbf{u}}$ and the shared ternary representation. For valid ligand atoms, we use the mean atom-type cross-entropy:

$$\mathcal{L}_{\text{atom}} = -\mathbb{E}\left[\frac{1}{N_{\mathcal{M}}}\sum_i \log \widehat{p}_{i,1}^{\,\text{a}}[a_i]\right]. \tag{66}$$

For bonds, one Gaussian noise vector is sampled per unordered atom pair and mirrored across $(i,j)$ and $(j,i)$, preserving symmetry throughout the flow. Because non-bonded pairs are substantially more frequent, the bond objective gives equal weight to bonded and non-bonded pairs within each molecule:

$$\mathcal{L}_{\text{bond}} = \frac{1}{2}\mathbb{E}\left[\text{CE}_{(i,j):b_{ij}\neq 0} + \text{CE}_{(i,j):b_{ij}=0}\right]. \tag{67}$$

**Riemannian CFM for Rigid Transformation**. Following (Bose et al., 2024), we predict $(R, \mathbf{t})$ within the flow matching framework, where the interpolation of orientations is performed along geodesics on $\mathrm{SO}(3)$, and translations are handled in Euclidean space.

For rotation, we employ geodesic interpolation on $\mathrm{SO}(3)$, which serves as the manifold counterpart of linear interpolation in Euclidean space and guarantees shortest-path evolution. Formally, given a random prior $\mathbf{R}_0 \sim U(\mathrm{SO}(3))$ and a target rotation $\mathbf{R}_1 \sim p(\mathbf{R} \mid \star)$, the conditional flow is:

$$\boldsymbol{\psi}_t(\mathbf{R}_0 \mid \mathbf{R}_1) = \exp_{\mathbf{R}_0}\big(t \cdot \log_{\mathbf{R}_0}(\mathbf{R}_1)\big), \tag{68}$$

where $\exp$ and $\log$ denote the Riemannian exponential and logarithm maps on $\mathrm{SO}(3)$, computable via Rodrigues' formula. The associated vector field is defined as:

$$u_t(\mathbf{R}_t \mid \mathbf{R}_0, \mathbf{R}_1) = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\psi}_t = \frac{\log_{\mathbf{R}_t}(\mathbf{R}_1)}{1-t}. \tag{69}$$

Then the CFM objective is formulated as:

$$\mathcal{L} = \mathbb{E}_{t\sim\mathcal{U}(0,1)}\left[\left\| v_\theta(\mathbf{R}_t, t, \star) - \frac{\log_{\mathbf{R}_t}(\mathbf{R}_1)}{1-t}\right\|_2^2\right]. \tag{70}$$

For translations, a Euclidean conditional flow is defined between $\mathbf{t}_0 \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$ and $\mathbf{t}_1 \sim p(\mathbf{t} \mid \star)$. We normalize its physical vector field using $s_{\text{trans}} = 50$ Å:

$$\widetilde{\mathbf{u}}_t^{\text{trans}} = \frac{\mathbf{t}_1 - \mathbf{t}_t}{(1-t)s_{\text{trans}}}. \tag{71}$$

**Training Losses**. The Coordinate block uses endpoint parameterization and minimizes the mean squared coordinate error over the $N_{\mathcal{M}}$ valid ligand atoms:

$$\mathcal{L}_{\text{coor}} = \mathbb{E}_t \left[ \frac{1}{N_{\mathcal{M}}} \sum_{i=1}^{N_{\mathcal{M}}} \|\widehat{\mathbf{x}}_{i,1} - \mathbf{x}_{i,1}\|_2^2 \right]. \tag{72}$$

The Rigid Transformation block directly regresses the rotational and normalized translational vector fields:

$$\mathcal{L}_{\text{rot}} = \mathbb{E}_t \left[ \|\boldsymbol{\omega}_t - \log_{\mathbf{R}_t}(\mathbf{R}_1)/(1-t)\|_2^2 \right], \tag{73}$$

$$\mathcal{L}_{\text{trans}} = \mathbb{E}_t \left[ \|\widetilde{\mathbf{v}}_t - (\mathbf{t}_1 - \mathbf{t}_t)/((1-t)s_{\text{trans}})\|_2^2 \right]. \tag{74}$$

We additionally supervise the assembled target coordinates to couple rotation and translation. Let $\widehat{\mathbf{X}}_{\mathcal{T},1} = \widehat{\mathbf{R}}_1 \mathbf{X}_{\mathcal{T}}^0 + \widehat{\mathbf{t}}_1$ and $\mathbf{X}_{\mathcal{T},1} = \mathbf{R}_1 \mathbf{X}_{\mathcal{T}}^0 + \mathbf{t}_1$. The pose-coordinate loss is

$$\mathcal{L}_{\text{pose}} = \mathbb{E}_t \left[ \frac{1}{N_{\mathcal{T}}} \sum_{i=1}^{N_{\mathcal{T}}} \text{SmoothL1}_\beta \Big( \widehat{\mathbf{X}}_{\mathcal{T},1,i} - \mathbf{X}_{\mathcal{T},1,i} \Big) \right], \qquad \beta = 2\,\text{Å}. \tag{75}$$

Together with the atom and bond losses above, the complete objective is

$$\mathcal{L}_{\text{flow}} = \mathcal{L}_{\text{trans}} + \mathcal{L}_{\text{rot}} + \mathcal{L}_{\text{atom}} + \mathcal{L}_{\text{bond}} + 5\mathcal{L}_{\text{coor}} + \mathcal{L}_{\text{pose}}. \tag{76}$$

# F Sample Algorithm

**Algorithm 1** Sampling with Interface-Conditioned Ternary Flow Network

**Input:** receptor $\mathcal{G}_{\mathcal{R}}$, target $\mathcal{G}_{\mathcal{T}^0}$, and grid $0 = t_0 < \cdots < t_N = 1$
$\hat{\mathcal{I}} = (\mathbf{Y}_{\mathcal{R}}, \mathbf{Y}_{\mathcal{T}}, \mathbf{\Sigma}_{\mathcal{R}}, \mathbf{\Sigma}_{\mathcal{T}}) \leftarrow \text{InterfaceModel}(\mathcal{G}_{\mathcal{R}}, \mathcal{G}_{\mathcal{T}^0})$ ▷ Estimate interface geometry
$(\mathbf{R}^*, \mathbf{t}^*) \leftarrow \text{KabschAlign}(\mathbf{Y}_{\mathcal{T}}, \mathbf{Y}_{\mathcal{R}})$ ▷ Construct a coarse pose condition
Sample $\mathbf{R}_0 \sim U(\text{SO}(3))$ and $\mathbf{t}_0 \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$
Sample $\mathbf{X}_{\mathcal{M},0} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and attribute logits $\mathbf{c}_{\mathcal{M},0}$
**for** $n = 0$ to $N-1$ **do**
  $\Delta t_n \leftarrow t_{n+1} - t_n, \quad \gamma_n \leftarrow \Delta t_n/(1-t_n)$
  Sample current atom and bond types $(\mathbf{a}_n, \mathbf{b}_n)$ from $\mathbf{c}_{\mathcal{M},n}$
  $(\boldsymbol{\omega}_n, \widetilde{\mathbf{v}}_n, \widehat{\mathbf{X}}_{\mathcal{M},1}, \widehat{\mathbf{c}}_{\mathcal{M},1}) \leftarrow \text{TernaryFlowNet}(\mathbf{R}_n, \mathbf{t}_n, \mathbf{X}_{\mathcal{M},n}, \mathbf{a}_n, \mathbf{b}_n, t_n \mid \star)$
  $\mathbf{R}_{n+1} \leftarrow \text{Proj}_{\text{SO}(3)}(\mathbf{R}_n \exp(\Delta t_n [\boldsymbol{\omega}_n \times]))$ ▷ Update rotation
  $\mathbf{t}_{n+1} \leftarrow \mathbf{t}_n + \Delta t_n s_{\text{trans}} \widetilde{\mathbf{v}}_n$ ▷ Update translation
  $\mathbf{X}_{\mathcal{M},n+1} \leftarrow \mathbf{X}_{\mathcal{M},n} + \gamma_n(\widehat{\mathbf{X}}_{\mathcal{M},1} - \mathbf{X}_{\mathcal{M},n})$ ▷ Update ligand coordinates
  $\mathbf{c}_{\mathcal{M},n+1} \leftarrow \mathbf{c}_{\mathcal{M},n} + \gamma_n(\widehat{\mathbf{c}}_{\mathcal{M},1} - \mathbf{c}_{\mathcal{M},n})$ ▷ Update atom and bond logits
**end for**
Decode $(\mathbf{a}_N, \mathbf{b}_N)$ from $\mathbf{c}_{\mathcal{M},N}$ and set $\mathbf{X}_{\mathcal{T},N} = \mathbf{R}_N \mathbf{X}_{\mathcal{T}^0} + \mathbf{t}_N$
**return** $\mathcal{C} = (\mathcal{R}, \mathcal{M}, \mathcal{T})$ defined by $\{\mathbf{R}_N, \mathbf{t}_N, \mathbf{X}_{\mathcal{M},N}, \mathbf{a}_N, \mathbf{b}_N\}$

Algorithm 1 summarizes the sampling procedure of TriGlue. Given two unbound proteins, the interface model predicts their interface geometry, from which Kabsch alignment derives a coarse pose condition without replacing the noisy target pose. Starting from independently sampled pose, coordinate, and attribute priors, the ternary flow network jointly evolves all components. The target rotation and translation follow the predicted rigid vector field, while ligand coordinates and attribute logits move toward their predicted clean endpoints. After $N$ steps, the final pose and molecular glue are decoded into the ternary complex.

# G PROOFS

**SE(3) Equivariance of Virtual Interface and Ellipsoid Parameters.**

**Theorem 1.** *The virtual interface points $\{\mathbf{y}_k\}_{k=1}^K$ and the ellipsoid parameters $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ defined in Sec. 3.2 are SE(3)-equivariant.*

*Proof.* Consider a rigid-body transformation $(\mathbf{R}, \mathbf{t}) \in \mathrm{SE}(3)$ applied to all residue coordinates:

$$\mathbf{z}_i' = \mathbf{R}\mathbf{z}_i + \mathbf{t}. \tag{77}$$

We first prove that the virtual interface points are equivariant. Since the attention weights are computed from EGNN invariant features, they remain unchanged under SE(3). Thus, we have:

$$\mathbf{y}_k' = \sum_i \alpha_i^k \mathbf{z}_i' = \sum_i \alpha_i^k (\mathbf{R}\mathbf{z}_i + \mathbf{t}) \tag{78}$$

$$= \mathbf{R}\sum_i \alpha_i^k \mathbf{z}_i + \mathbf{t} = \mathbf{R}\mathbf{y}_k + \mathbf{t}. \tag{79}$$

Hence $\mathbf{y}_k$ are SE(3)-equivariant. Then, it is not hard to obtain that:

$$\hat{\boldsymbol{\mu}}' = \frac{1}{K}\sum_k \mathbf{y}_k' = \frac{1}{K}\sum_k (\mathbf{R}\mathbf{y}_k + \mathbf{t}) = \mathbf{R}\hat{\boldsymbol{\mu}} + \mathbf{t}, \tag{80}$$

and the centered coordinates satisfy:

$$\mathbf{y}_k' - \hat{\boldsymbol{\mu}}' = \mathbf{R}(\mathbf{y}_k - \hat{\boldsymbol{\mu}}), \tag{81}$$

which yields:

$$\hat{\boldsymbol{\Sigma}}' = \frac{1}{K}\sum_k (\mathbf{y}_k' - \hat{\boldsymbol{\mu}}')(\mathbf{y}_k' - \hat{\boldsymbol{\mu}}')^\top \tag{82}$$

$$= \mathbf{R}\hat{\boldsymbol{\Sigma}}\mathbf{R}^\top. \tag{83}$$

Therefore, $(\mathbf{y}_k, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ transform equivariantly under SE(3). □

**SE(3) Equivariance of the Rigid Transformation Block.**

**Theorem 2.** *The Rigid Transformation block $\Phi_{\mathrm{RT}}$ defined in Appendix D.3 is SE(3)-equivariant with respect to the receptor reference frame. Specifically, under a global transformation $(\mathbf{Q}, \mathbf{s}) \in \mathrm{SE}(3)$ of all world-frame quantities, its world-frame outputs satisfy*

$$\mathbf{v}_t' = \mathbf{Q}\mathbf{v}_t, \qquad (\boldsymbol{\omega}_t^{\mathrm{world}})' = \mathbf{Q}\boldsymbol{\omega}_t^{\mathrm{world}}, \tag{84}$$

*while the right-trivialized body-frame angular field satisfies $\boldsymbol{\omega}_t' = \boldsymbol{\omega}_t$.*

*Proof.* The unbound target $\mathcal{T}^0$ is a rigid template expressed in its fixed body frame, whereas the receptor, assembled target, ligand, and receptor-side interface are expressed in the receptor world frame. Thus, the global transformation changes the world frame but does not transform the internal coordinates $\mathbf{X}_{\mathcal{T}}^0$ or $\mathbf{Y}_{\mathcal{T}}$ of the target template. Every world-frame point obeys $\mathbf{x}' = \mathbf{Q}\mathbf{x} + \mathbf{s}$. Since $\mathbf{X}_{\mathcal{T},t} = \mathbf{R}_t\mathbf{X}_{\mathcal{T}}^0 + \mathbf{t}_t$, the current and Kabsch poses consequently transform as

$$\mathbf{R}_t' = \mathbf{Q}\mathbf{R}_t, \quad \mathbf{t}_t' = \mathbf{Q}\mathbf{t}_t + \mathbf{s}, \qquad (\mathbf{R}^*)' = \mathbf{Q}\mathbf{R}^*, \quad (\mathbf{t}^*)' = \mathbf{Q}\mathbf{t}^* + \mathbf{s}. \tag{85}$$

The EGNN features, pairwise distances, contact weights, and timestep features are invariant. Translation cancels in every displacement, while rotations preserve inner products and satisfy

$$(\mathbf{Q}\mathbf{a}) \times (\mathbf{Q}\mathbf{b}) = \mathbf{Q}(\mathbf{a} \times \mathbf{b}). \tag{86}$$

Consequently, every vector in $\mathcal{B}_{\mathrm{trans}}$ and $\mathcal{B}_{\mathrm{rot}}$, as well as the local force and torque, is multiplied by $\mathbf{Q}$. The Kabsch rotation basis follows the same rule because

$$\boldsymbol{\rho}_K' = \mathbf{R}_t'\big(\log((\mathbf{R}_t')^\top(\mathbf{R}^*)')\big)^\vee = \mathbf{Q}\boldsymbol{\rho}_K. \tag{87}$$

Moreover, norms, dot products, and covariance discrepancies are invariant under rotation; hence $\mathbf{z}_t$ and all learned coefficients remain unchanged. Since radial normalization also commutes with rotation, Eq. (52) yields

$$\mathbf{v}'_t = \mathbf{Q}\mathbf{v}_t, \qquad (\boldsymbol{\omega}_t^{\text{world}})' = \mathbf{Q}\boldsymbol{\omega}_t^{\text{world}}. \tag{88}$$

Finally,

$$\boldsymbol{\omega}'_t = (\mathbf{R}'_t)^\top (\boldsymbol{\omega}_t^{\text{world}})' = \mathbf{R}_t^\top \boldsymbol{\omega}_t^{\text{world}} = \boldsymbol{\omega}_t. \tag{89}$$

Therefore the endpoint update satisfies $\widehat{\mathbf{R}}'_1 = \mathbf{Q}\widehat{\mathbf{R}}_1$ and $\widehat{\mathbf{t}}'_1 = \mathbf{Q}\widehat{\mathbf{t}}_1 + \mathbf{s}$, proving SE(3) equivariance of $\Phi_{\text{RT}}$. □

# H Experimental Details

## H.1 Dataset Details

We use the molecular-glue-induced ternary-complex split derived from TernaryDB[1] (Xue et al., 2025). Each sample contains two protein components, including a target protein and an E3 ligase, and one small-molecule ligand extracted from an experimentally resolved ternary complex. Protein lengths are calculated based on the number of resolved C$\alpha$ residues in the curated protein PDB files. Ligand descriptors are computed from the per-complex SDF files using RDKit. QED denotes the quantitative estimate of drug-likeness. Detailed statistics are provided in Tab. 4. For numerical descriptors, we report the median, with the minimum and maximum values.

Table 4: Summary statistics of the training and held-out molecular-glue test splits. Numerical descriptors are reported as **Median (Minimum, Maximum)**.

| Component | Statistic | Training set | Test set |
|---|---|---|---|
| *Complex* | Number of ternary complexes | 21,481 | 94 |
| | **Sequence-pair overlap with training set** | – | **0/94** |
| *Protein* | Length of E3 ligase | 321 (3, 1765) | 227 (107, 826) |
| | Length of target protein | 125 (3, 1759) | 76 (2, 512) |
| *Molecular Glue* | Number of heavy atoms | 32 (2, 100) | 25 (10, 69) |
| | Molecular weight (Da) | 506.20 (26.04, 1464.07) | 385.85 (136.15, 980.31) |
| | Number of rings | 1 (0, 24) | 3 (0, 13) |
| | Number of aromatic rings | 0 (0, 7) | 0 (0, 5) |
| | QED | 0.24 (0.02, 0.93) | 0.54 (0.12, 0.88) |

## H.2 Baseline Details

**Baselines**. Our framework jointly performs *de novo* molecular glue generation and reconstructs the unbound target protein into the ternary-complex coordinate frame. However, no existing method evaluates both aspects simultaneously. Therefore, we select and compare representative baselines separately for each task. Overall, these baselines span a broad spectrum of representative methods:

(i) *Molecular design method.* To assess the quality of generated molecular glues, three structure-based drug design methods are considered: **AR**[2] (Luo et al., 2021), **TargetDiff**[3] (Guan et al., 2023a), and **PocketXMol**[4] (Peng et al., 2026), which perform *de novo* small-molecule generation conditioned on protein binding pockets.

(ii) *Ternary complex structure prediction method.* **DeepTernary**[5] (Xue et al., 2025) is the first model designed for ternary complex structure prediction. We adopt it to dock the unbound protein back into the ternary complex, thereby evaluating the ability to model the complex structure.

[1] https://doi.org/10.5281/zenodo.15514874
[2] https://github.com/luost26/3D-Generative-SBDD
[3] https://github.com/guanjq/targetdiff
[4] https://github.com/pengxingang/PocketXMol
[5] https://github.com/youqingxiaozhua/DeepTernary

(iii) *Protein-protein docking method.* To further evaluate the rigid-body docking capability, we include two representative protein-protein docking approaches: **EquiDock**[6] (Ganea et al., 2022) and **DiffDock-PP**[7] (Ketata et al., 2023), which predict the relative pose of interacting proteins from their unbound structures.

All of the baselines are implemented with the official code.

### H.3 Training Details

The proposed framework consists of two stages: interface estimation and interface-conditioned ternary flow generation.

**Interface Estimation Module**. The interface estimation model was trained for 1200 epochs with a batch size of 8 using distributed training on 8 GPUs. It uses a four-layer EGNN with a hidden dimension of 128, 16 nearest neighbors, and $K = 50$ virtual interface points. We optimized it with AdamW using an initial learning rate of $5 \times 10^{-4}$ and cosine annealing.

**Interface-Conditioned Ternary Flow Network**. The ternary flow model uses a four-layer unified EGNN with a hidden dimension of 128 and was trained for 800 epochs. We used a batch size of 12 per GPU on 8 GPUs, giving an effective batch size of 96, with BF16 mixed precision. The pretrained interface estimation module was frozen during this stage. We optimized the network with Adam and linearly decayed the learning rate from $5 \times 10^{-4}$ to $3 \times 10^{-4}$. The weights for the translation, rotation, atom-type, bond-type, coordinate, and pose-coordinate losses were 1, 1, 1, 1, 5, and 1, respectively. The maximum gradient norm was set to 10.

Translations use Euclidean conditional flow matching with Gaussian noise $\sigma = 1.0$ and a normalization scale of 50 Å, while rotations use Riemannian flow matching on $\mathrm{SO}(3)$. Atom and bond types follow simplex paths with scale $K = 5.0$, using 25 atom categories and 6 bond categories. We used 100 integration steps for the reported evaluations.

**Dataset Split**. The ternary complex dataset was divided into training and validation subsets using a 99%/1% split. Unless otherwise specified, all experiments used a random seed of 42.

Training the interface model takes approximately 2 days, training the ternary flow network requires about 5 days, and sampling a single complex takes around 40 seconds.

### H.4 Evaluation Metric Details

We evaluate the models from the following two perspectives:

(i) **Molecular generation**. We assess the generated ligands with respect to binding and physicochemical properties. Binding affinity is estimated using AutoDock Vina (Trott & Olson, 2010), including the **Vina score**, the minimized Vina score (**Vina min**) after local relaxation, and the redocking Vina score (**Vina dock**) that reflects the best achievable binding affinity. The high-affinity ratio (**Aff**) is defined as the proportion of generated molecules whose Vina dock surpasses that of the reference ligand for each target. To further examine chemical quality, we evaluate standard molecular property metrics, including quantitative estimation of drug-likeness (**QED**) and the synthetic accessibility (**SA**) score.

(ii) **Ternary-complex reconstruction**. We evaluate the structural accuracy of the generated complex from both ligand and protein perspectives. Ligand pose accuracy is measured using the root-mean-square deviation (**RMSD**) between predicted and reference ligand coordinates. Protein docking accuracy is assessed using the **DockQ** score (Basu & Wallner, 2016), and we additionally report the success rate with DockQ greater than 0.23, which is commonly regarded as indicating acceptable docking quality.

**Root Mean Square Deviation (RMSD)**. Ligand pose quality is assessed using the RMSD between the predicted ligand coordinates and the reference (ground truth) coordinates.

[6] https://github.com/octavian-ganea/equidock_public
[7] https://github.com/ketatam/DiffDock-PP

Let $\mathbf{x}_i \in \mathbb{R}^3$ denote the predicted coordinate of the $i$-th ligand atom and $\mathbf{x}_i^* \in \mathbb{R}^3$ denote the corresponding reference coordinate. Given $N$ ligand atoms, RMSD is defined as:

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i - \mathbf{x}_i^*\|_2^2}. \tag{90}$$

Lower RMSD values indicate higher agreement between the predicted ligand pose and the reference structure.

**AutoDock Vina-based Binding Metrics.** To evaluate the predicted protein-ligand binding quality of generated molecular glues, we report three AutoDock Vina-based scores: Vina score, Vina min, and Vina dock. Let $P$ denote the protein binding environment, $M$ denote a generated molecule, and $\hat{\mathbf{x}}$ denote its generated 3D conformation. We write $E_{\mathrm{Vina}}(P, M, \mathbf{x})$ as the AutoDock Vina scoring function, which approximates the protein-ligand binding free energy in kcal/mol. More negative values indicate stronger predicted binding affinity.

The **Vina score** is computed on the generated ligand pose without further pose optimization:

$$\mathrm{VinaScore}(M) = E_{\mathrm{Vina}}(P, M, \hat{\mathbf{x}}). \tag{91}$$

This metric evaluates whether the model directly generates a chemically plausible and energetically favourable binding pose.

**Vina min** further performs local energy minimization around the generated pose before scoring:

$$\mathrm{VinaMin}(M) = \min_{\mathbf{x}\in\mathcal{N}(\hat{\mathbf{x}})} E_{\mathrm{Vina}}(P, M, \mathbf{x}), \tag{92}$$

where $\mathcal{N}(\hat{\mathbf{x}})$ denotes the local conformational neighbourhood explored during minimization. Compared with the Vina score, Vina min is less sensitive to small local geometric errors in the generated conformation.

**Vina dock** evaluates the molecule after a full re-docking procedure:

$$\mathrm{VinaDock}(M) = \min_{\mathbf{x}\in\Omega(P,M)} E_{\mathrm{Vina}}(P, M, \mathbf{x}), \tag{93}$$

where $\Omega(P, M)$ denotes the docking search space of possible ligand poses in the binding region. This metric estimates the best binding score achievable by the generated molecule under the docking protocol, rather than only evaluating the initially generated pose. For all three Vina-based metrics, lower values are better.

**Quantitative Estimate of Drug-likeness (QED).** Drug-likeness is evaluated using the Quantitative Estimate of Drug-likeness (QED), which summarizes how similar a molecule is to known drug-like compounds. QED combines multiple molecular descriptors, including molecular weight, lipophilicity, topological polar surface area, hydrogen bond donors and acceptors, aromatic rings, rotatable bonds, and structural alerts. Given descriptor-specific desirability functions $d_k(M) \in [0, 1]$ and weights $w_k$, QED is computed as a weighted geometric mean:

$$\mathrm{QED}(M) = \exp\left(\frac{\sum_{k=1}^{K} w_k \log d_k(M)}{\sum_{k=1}^{K} w_k}\right). \tag{94}$$

QED ranges from 0 to 1, where larger values indicate more favourable drug-like properties.

**Synthetic Accessibility (SA).** Synthetic accessibility measures how easy a generated molecule is expected to be synthesized. The commonly used raw synthetic accessibility score is based on fragment contributions and molecular complexity penalties:

$$\mathrm{SAscore}_{\mathrm{raw}}(M) = S_{\mathrm{frag}}(M) + S_{\mathrm{complexity}}(M) + S_{\mathrm{correction}}(M). \tag{95}$$

Here, $S_{\mathrm{frag}}(M)$ reflects whether the molecule contains fragments frequently observed in known synthesized compounds, while $S_{\mathrm{complexity}}(M)$ penalizes structural features that make synthesis difficult, such as large rings, complex ring systems, stereochemical complexity, and large molecular size.

Since the raw SA score is usually defined such that lower values indicate easier synthesis, while our table reports SA with a larger-is-better direction, we use a normalized synthesizability score:

$$\mathrm{SA}(M) = \mathrm{clip}\left(\frac{10 - \mathrm{SAscore}_{\mathrm{raw}}(M)}{9}, 0, 1\right). \tag{96}$$

Under this convention, larger SA values indicate molecules that are predicted to be easier to synthesize.

**High Affinity Rate (Aff).** Following common evaluation protocols in target-aware molecule generation, Aff measures the proportion of generated molecules whose predicted binding affinity is stronger than that of the reference ligand for the same target. Let $s_{t,j}^{\mathrm{dock}}$ denote the Vina dock score of the $j$-th generated molecule for target $t$, and let $s_t^{\mathrm{ref}}$ denote the corresponding score of the reference ligand. Since lower Vina scores indicate stronger predicted binding, Aff is defined as:

$$\mathrm{Aff} = \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}}\frac{1}{K_t}\sum_{j=1}^{K_t}\mathbb{I}\left[s_{t,j}^{\mathrm{dock}} < s_t^{\mathrm{ref}}\right], \tag{97}$$

where $\mathcal{T}$ is the set of test targets, $K_t$ is the number of generated molecules for target $t$, and $\mathbb{I}[\cdot]$ is the indicator function. Higher Aff values indicate that a larger fraction of generated molecules are predicted to bind more strongly than the reference ligand.

**DockQ.** For protein–protein docking, model quality is evaluated using DockQ, a continuous score in the range $[0, 1]$ that summarizes interface accuracy by combining three complementary aspects. (i) The fraction of native contacts ($F_{\mathrm{nat}}$) measures how many residue–residue contacts observed in the reference complex are correctly reproduced in the prediction, reflecting interface contact recovery. (ii) The interface RMSD (iRMSD) quantifies the structural deviation of backbone atoms for residues located at the binding interface after superposition of the complexes, capturing local interface accuracy. (iii) The ligand RMSD (LRMSD) measures the backbone deviation of the ligand protein after aligning the receptor, reflecting the global placement of the binding partner. These three components are smoothly combined into a single score:

$$\mathrm{DockQ} = \frac{1}{3}\left(F_{\mathrm{nat}} + \frac{1}{1+\left(\frac{\mathrm{iRMSD}}{1.5}\right)^2} + \frac{1}{1+\left(\frac{\mathrm{LRMSD}}{8.5}\right)^2}\right). \tag{98}$$

### H.5 Ellipsoid Interface Prediction Metric Details

Each interface is represented as a Gaussian ellipsoid $\mathcal{N}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^3$ denotes the ellipsoid center and $\Sigma \in \mathbb{R}^{3\times 3}$ is a symmetric positive definite (SPD) matrix describing its orientation and spatial extent. Given a prediction $(\mu_p, \Sigma_p)$ and reference $(\mu_r, \Sigma_r)$, we evaluate complementary geometric and distributional discrepancies using the following metrics.

**Center-L2**. This metric measures the Euclidean distance between predicted and reference ellipsoid centers:

$$d_{\mathrm{center}} = \|\mu_p - \mu_r\|_2. \tag{99}$$

**LogDet Error**. Differences in ellipsoid scale are quantified by comparing the log-determinants of the covariance matrices:

$$d_{\log\det} = \left|\log\det(\Sigma_p) - \log\det(\Sigma_r)\right|. \tag{100}$$

**Wasserstein-2 distance (W2 Dist)**. This geometry-aware distance jointly evaluates discrepancies in both the ellipsoid center and shape:

$$W_2^2 = \|\mu_p - \mu_r\|_2^2 + \mathrm{Tr}\left(\Sigma_p + \Sigma_r - 2\left(\Sigma_r^{1/2}\Sigma_p\Sigma_r^{1/2}\right)^{1/2}\right), \tag{101}$$

and we report $W_2 = \sqrt{W_2^2}$.

**Jensen–Shannon divergence (JSD)**. Distributional similarity between the predicted and reference Gaussian densities is assessed using JSD:

$$\mathrm{JSD}(P\|Q) = \frac{1}{2}\mathrm{KL}(P\|M) + \frac{1}{2}\mathrm{KL}(Q\|M), \quad M = \frac{1}{2}(P+Q), \tag{102}$$

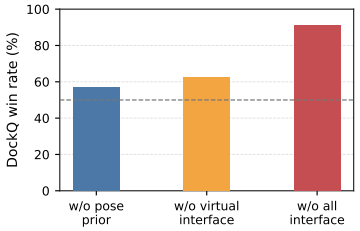



Figure 5: Paired DockQ win rate of TriGlue over each interface-ablated variant. The dashed line marks a 50% win rate.

where $P = \mathcal{N}(\mu_r, \Sigma_r)$ and $Q = \mathcal{N}(\mu_p, \Sigma_p)$. Because $M$ is a Gaussian mixture rather than a single Gaussian, JSD is estimated via Monte Carlo:

$$\widehat{\mathrm{JSD}} = \frac{1}{2}\left[\frac{1}{N}\sum_{x_i \sim P} \log \frac{p(x_i)}{m(x_i)} + \frac{1}{N}\sum_{y_j \sim Q} \log \frac{q(y_j)}{m(y_j)}\right], \tag{103}$$

where $m(\cdot) = \frac{1}{2}\big(p(\cdot) + q(\cdot)\big)$.

### H.6 TriConf Definition

For a protein pair, let $\mathcal{E}_i^{(0)}$ denote the interface ellipsoid predicted for protein $i \in \{1, 2\}$ from the original input. We generate $M = 32$ perturbed inputs using 0.2 Å coordinate noise and 10% residue masking, and denote their predictions by $\mathcal{E}_i^{(m)}$. The interface deviation under perturbation $m$ and the resulting confidence score are defined as:

$$\begin{aligned} d_m &= \frac{1}{2}\sum_{i=1}^{2} W_2\Big(\mathcal{E}_i^{(m)}, \mathcal{E}_i^{(0)}\Big), \quad m = 1, \ldots, M, \\ \mathrm{TriConf} &= \left(1 + \frac{\mathrm{Med}(d_1, \ldots, d_M)}{1\,\text{Å}}\right)^{-1}. \end{aligned} \tag{104}$$

TriConf lies in $(0, 1]$, with higher values corresponding to smaller perturbation-induced interface deviations. Its computation uses only the predicted interface ellipsoids and therefore does not require the ground-truth ternary complex.

## I Additional Experiment Results

### I.1 Interface Ablation Details

We isolate the two routes through which the predicted interface conditions the Rigid Transformation block (Appendix D.3): the Kabsch pose prior $(\mathbf{R}^*, \mathbf{t}^*)$ and the virtual-interface geometry $(\mathbf{Y}_{\mathcal{R}}, \mathbf{Y}_{\tilde{\mathcal{T}}}, \mathbf{\Sigma}_{\mathcal{R}}, \mathbf{\Sigma}_{\tilde{\mathcal{T}}})$. All variants use the same epoch-800 checkpoint and matched sampling noise, so differences arise only from the conditions available to $\Phi_{\mathrm{RT}}$. Specifically, **w/o pose prior** masks the Kabsch-derived bases and invariants, **w/o virtual interface** masks the displacement, torque, and covariance cues derived from the virtual interfaces, and **w/o all interface** masks both routes. The ligand-conditioned ternary representations and local force–torque field are retained in every variant.

We sample 20 trajectories with 100 integration steps and a fixed random seed of 42 for each test complex. DockQ is available for 91 complexes, and the maximum DockQ over the 20 trajectories is used for each paired comparison. The bars in Fig. 4b report the mean across complexes.

As shown in Fig. 5, TriGlue outperforms the variants without the pose prior, virtual interface, and all interface conditions on $57.1\%$, $62.6\%$, and $91.2\%$ of paired complexes, respectively. Removing the virtual interface reduces mean DockQ by $0.062$ ($p = 0.003$, paired Wilcoxon test), while removing the pose prior produces a smaller reduction of $0.027$ ($p = 0.172$). Removing both conditioning routes

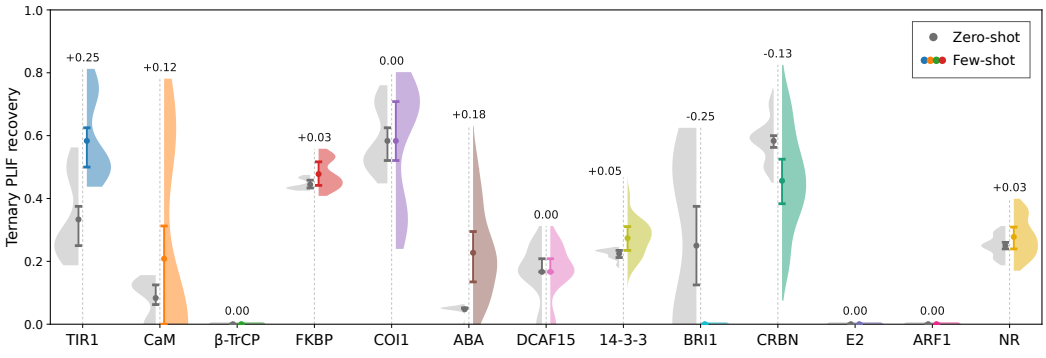



Figure 6: Family-wise ternary PLIF recovery before and after few-shot adaptation. Gray and colored distributions denote zero-shot and few-shot results, respectively; annotations report changes in mean recovery.

causes the largest degradation: mean DockQ decreases by $0.134$ ($p = 1.4 \times 10^{-13}$), and mean target LRMSD increases from $18.46$ Å to $27.95$ Å. These results indicate that the detailed virtual-interface geometry provides the stronger individual signal, while the two interface-conditioning routes together are important for robust ternary-pose recovery.

### I.2 Ternary Interaction Recovery after Family Adaptation

We additionally evaluate whether family-level fine-tuning improves ligand interactions with both proteins using a ternary protein–ligand interaction fingerprint (PLIF). Each fingerprint feature records the protein side, residue identity, and interaction type detected by ProLIF. Let $\mathcal{F}_{\text{pred}}$ and $\mathcal{F}_{\text{ref}}$ denote the union of ligand–receptor and ligand–target features in the generated and native complexes, respectively. We define ternary PLIF recovery as:

$$\text{PLIF}_{\text{rec}} = \frac{|\mathcal{F}_{\text{pred}} \cap \mathcal{F}_{\text{ref}}|}{|\mathcal{F}_{\text{ref}}|}. \tag{105}$$

To isolate geometric interaction recovery, we use the native ligand topology with the generated ligand coordinates and protein pose, selecting the minimum-ligand-RMSD trajectory for both settings. Across 886 paired query evaluations, family adaptation increases mean PLIF recovery from 0.206 to 0.274 ($\Delta = 0.068$, Wilcoxon $p = 6.3 \times 10^{-9}$). Figure 6 shows the family-wise changes.

### I.3 Robustness to Predicted Apo Structures

In prospective molecular-glue design, an experimentally resolved receptor structure may not be available. We therefore evaluate whether TriGlue, which is trained with experimental structures, remains effective when its receptor input is replaced by a structure predicted from sequence. Among the 92 test complexes, we identify 75 unique receptor sequences and predict each receptor independently using AlphaFold3 (Abramson et al., 2024). Each AlphaFold3 job contains only the receptor sequence. For 71 sequences, we reuse the previously computed unpaired MSA and template features; the remaining four are processed with the standard AlphaFold3 data pipeline. We explicitly remove all paired-MSA features, preventing information about the target protein or ligand from entering this receptor-only control.

To isolate the effect of receptor structure quality, we rigidly align each predicted receptor to the coordinate frame of its corresponding experimental receptor using C$\alpha$ atoms and replace only the receptor coordinates in the original test input. The experimentally determined target structure, reference complex, and all other input fields are kept unchanged. We evaluate the same pretrained checkpoint without additional fine-tuning and use the same sampling protocol for both receptor inputs, drawing 100 trajectories per complex. For each complex, we report the highest DockQ and lowest ligand RMSD among these trajectories; valid DockQ values are available for 91 of the 92 paired complexes.

Table 5: Robustness to an AlphaFold3-predicted apo receptor. Results are reported as mean ± standard deviation over the best of 100 sampled trajectories per complex; success denotes DockQ $> 0.23$.

| Receptor input | DockQ ↑ | Success (%) ↑ | RMSD (Å) ↓ |
|---|---|---|---|
| Experimental | $0.309 \pm 0.220$ | 52.7 | $4.58 \pm 1.26$ |
| AlphaFold3 | $0.305 \pm 0.232$ | 45.1 | $4.66 \pm 1.31$ |

As shown in Tab. 5, replacing the experimental receptor with its AlphaFold3 prediction changes the mean best DockQ by only $-0.004$ (0.309 to 0.305) and increases the mean best ligand RMSD by 0.08 Å (4.58 to 4.66 Å). The acceptable-docking rate decreases from 52.7% to 45.1%, indicating that structural prediction errors can affect complexes close to the DockQ success threshold. Nevertheless, the small changes in the continuous metrics show that TriGlue largely preserves its reconstruction quality with receptor-only predicted structures, supporting its applicability when experimentally resolved receptor structures are unavailable.

### I.4 Interface Representation Analysis

To further analyze the representations learned by the interface model, we cluster the ternary complexes in the test set by structural and functional similarity, yielding several groups represented by one or two complexes (e.g., `2O98_A_P_FSC`). We select six representative groups and encode the two proteins in each complex with the interface model. As shown in Fig. 7, a t-SNE visualization (van der Maaten & Hinton, 2008) of the resulting hidden states forms compact and well-separated clusters that closely align with the structural grouping of the complexes. This result indicates that the model captures transferable, high-level interface geometry and interaction patterns rather than merely fitting individual samples.

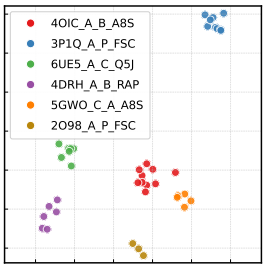



Figure 7: t-SNE visualization of the interface model's hidden states for six representative groups.

## J Family-Wise Case Studies of Ternary-Complex Design

Figure 8 presents representative successful and challenging cases across molecular-glue families. The most consistent behavior is observed for the 14-3-3–client family: seven complexes involving different clients and glues achieve DockQ values between 0.60 and 0.80 while retaining favorable QED and Vina Dock scores. This consistency indicates that TriGlue captures a transferable family-level interaction geometry rather than memorizing a single complex. The successful cases further extend to structurally distinct systems, including E2–ubiquitin, SCF/COI1–JAZ, nuclear receptor–corepressor, SCF/$\beta$-TrCP–$\beta$-catenin, SCF/TIR1–Aux/IAA, ABA receptor–PP2C, DCAF15–RBM39, and aldolase–peptide. In particular, the two SCF/COI1–JAZ examples perform favorably on both the protein-docking and molecular-quality axes, showing that the model's generalization extends beyond the dominant 14-3-3 family.

The final row provides a compact view of selected challenging systems. The CRBN–neosubstrate examples retain favorable molecular scores despite lower DockQ, indicating that molecular quality and protein-docking quality need not vary together. The contrast between `5MN0` and `4OIC` within the ABA receptor–PP2C family further illustrates instance-level variation within a family. These

examples are qualitative illustrations rather than a ranking of family difficulty, and motivate further study of family- and complex-specific interface geometry.

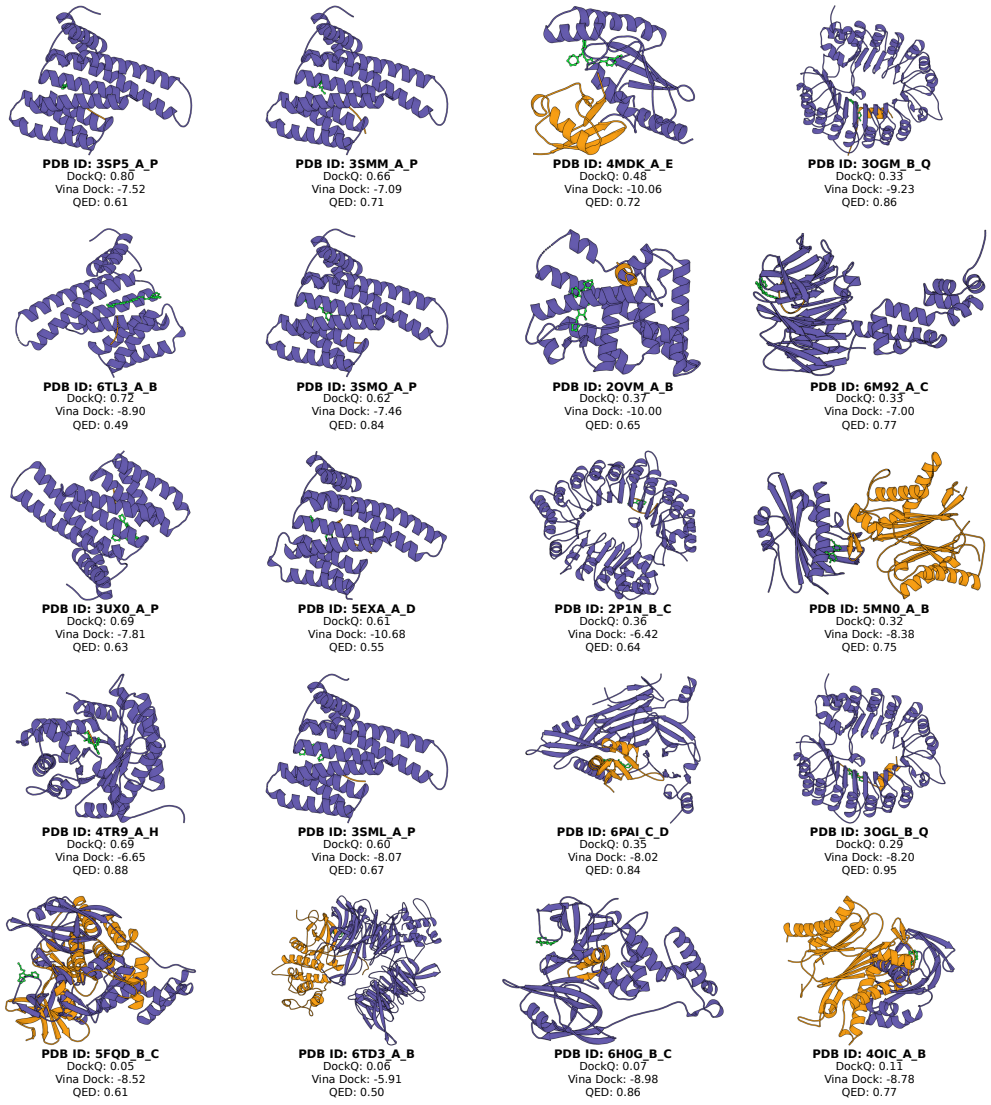



Figure 8: Family-wise case studies of TriGlue designs. The first four rows show representative successful examples across diverse families, including consistent reconstruction within the 14-3-3–client family and transfer to distinct protein systems. The final row illustrates selected challenging cases and the variation observed across and within families. These examples are qualitative and do not constitute a ranking of family difficulty. Receptor, target, and generated glue are shown in purple, orange, and green, respectively.