MULTIMODAL DIFFUSION OVER SEQUENCE AND $SE(3)$ MANIFOLDS FOR PROTEIN-PROTEIN COMPLEX DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Designing high-affinity target-specific protein binders is critical to modern therapeutics. Existing protein generation methods are unsatisfactory on this task because many of their designs fail to achieve joint geometric and physicochemical complementarity at the binding interfaces. To address this, we propose PPFrame, a physicochemically enhanced multimodal diffusion method for protein-protein complex design. The core idea is to co-design a target-conditional protein sequence and structure by jointly diffusing over the discrete sequence space and the continuous $SE(3)$ structural manifold. We also introduce IPAFormer to serve as the core network inside the diffusion. IPAFormer's architectural insight is its interleaving of geometry-based invariant point attention and global sequence self-attention. We curate PPIBench, a dataset of 602,364 protein-protein complexes across diverse target scaffolds from the Protein Data Bank (PDB). Our PPFrame is first pretrained on PPIBench and then finetuned on two important real-world tasks: mini-binder design and antibody design. PPFrame consistently outperforms the best baselines across the three tasks, improving design success rates by 17.8%, 8.0%, and 1.9%, respectively. Notably, PPFrame generates a candidate broadly neutralizing antibody (bNAb) against HIV-1 with an ipTM of 0.81 and a binding energy of -52.1, highlighting its strong utility for real-world biotherapeutic discovery.

1 INTRODUCTION

The *de novo* design of high-affinity, target-specific protein binders is foundational to modern biomedicine, driving transformative advances in therapeutics (Nelson et al., 2010), diagnostics (Brennan et al., 2010), and vaccines (Correia et al., 2014; Sesterhenn et al., 2020). Central to molecular recognition is the requirement for strict geometric and physicochemical complementarity across the binding interface (Chothia & Janin, 1975; Conte et al., 1999; Gainza et al., 2023). The complementary interface is governed by core energetic drivers: van der Waals forces optimize spatial packing, while hydrophobic and electrostatic interactions enforce physicochemical compatibility. Although traditional methods have exploited these energetic principles to engage pre-selected surface patches (Chevalier et al., 2017; Silva et al., 2019), they frequently fail when target proteins lack well-defined binding pockets, thereby restricting their scope to a narrow range of target topologies.

Recently, rapid advances in deep generative modeling have significantly expanded binder design capabilities (Evans et al., 2021; Cao et al., 2022; Bennett et al., 2023; Zambaldi et al., 2024; Stark et al., 2025; Sappington et al., 2026). These approaches have made it possible to design binders across diverse protein targets. In specialized domains, such as flexible peptides (Vázquez Torres et al., 2024) and intrinsically disordered proteins (Wu et al., 2024), deep generative models can yield potent binders without requiring extensive experimental optimization. Despite these breakthroughs, the overall success rate remains low due to two primary failure modes: the designed structure may fail to bind target interfaces due to the physicochemical or geometric non-complementarity, and the designed sequence may not fold into the intended structure. Crucially, most existing frameworks decouple sequence and structure generation or specialize in particular interaction types, leaving challenging, pocketless targets still intractable.

To address these challenges, we present PPFrame, a physicochemically enhanced multimodal diffusion method to co-design binder protein sequence and structure for an arbitrary target. In our framework, proteins are represented as rigid-body frames, coupling each residue's spatial backbone

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.

coordinates with its local physicochemical properties. PPFrame employs a unified multimodal diffusion process that operates jointly over discrete amino acid sequence space and the continuous $SE(3)$ structural manifold, progressively adding noise to train a joint generative model. The generative model is powered by IPAFormer, an architecture that interleaves invariant point attention (Jumper et al., 2021) with sequence self-attention (Vaswani et al., 2017) to simultaneously capture spatial geometric constraints and amino acid dependencies. By tightly integrating sequence, structure, and physicochemical properties, PPFrame enforces binding interface complementarity and high sequence-structure consistency, effectively mitigating key failure modes in binder design to drive higher success rates.

Concretely, our contributions are listed as follows:

- We propose PPFrame, a multimodal diffusion method building upon IPAFormer to co-design the binder protein sequences and structures for arbitrary protein targets.
- We create PPIBench, a general protein-protein complex dataset across diverse target geometries comprising 602,364 complexes curated from PDB (Berman et al., 2000).
- Our model pretrained on PPIBench outperforms strong baselines, demonstrating a sequence design success rate of 47.8%, structure binding energy of -67.8 and a co-design consistency of 65.2%. The main code is available at `https://anonymous.4open.science/r/PPFrame-3A1E` and the designed candidates are provided in the supplementary material.
- We finetune the pretrained PPFrame on two important real-world applications: target-protein mini-binder complex design and antigen-antibody complex design. The finetuned models achieve design success rates of 40.7% and 18.7%, respectively, demonstrating the framework's effectiveness in addressing key therapeutic challenges.

2 Related Work

Protein-Protein Complex Design. Designing target-specific protein binders with high affinity has been extensively explored using a wide range of approaches. Conventional experimental approaches rely primarily on directed evolution of natural protein scaffolds (Vazquez-Lombardi et al., 2015) or high-throughput antibody screening from synthetic libraries (Chao et al., 2006; Hackel et al., 2008). Although powerful, these methods require considerable experimental effort and do not provide substantial control over the properties of the binder proteins. Traditional computational methods have sought to model physicochemical properties to accelerate the design process (Chevalier et al., 2017; Fleishman et al., 2011). However, their reliance on pre-defined hotspot residues heavily restricts their scope. Recent advances in deep learning have significantly improved the prediction and design of protein-protein interactions (Evans et al., 2021; Humphreys et al., 2021; Bryant et al., 2022; Pacesa et al., 2024; Didi et al., 2026), enabling binder design without prior hotspot specification (Cao et al., 2022; Bennett et al., 2023; Chen et al., 2025). Nevertheless, the overall success rate remains low due to the difficulty of simultaneously satisfying the strict geometric and physicochemical complementarity essential for stable physical binding interfaces (Berger et al., 2024; Zambaldi et al., 2024).

Diffusion Models. Diffusion models (Sohl-Dickstein et al., 2015) have established themselves as a powerful class of generative frameworks. Ho et al. (2020) introduce denoising diffusion probabilistic models, formalizing their connection to score-based generative modeling (Song & Ermon, 2019). Diffusion models have achieved remarkable success in computational protein design. Early structural approaches focus on backbone design, progressing from C_α trace generation (Trippe et al., 2022) to full rigid-body $SE(3)$ backbone modeling (Watson et al., 2023; Yim et al., 2023). In parallel, discrete diffusion formulations are introduced to tackle protein sequence generation (Alamdari et al., 2023; Wang et al., 2024). More recent efforts have explored joint sequence-structure co-design through multimodal diffusion (Song et al., 2025; Wang et al., 2025). However, the application of diffusion models for the sequence-structure co-design of protein complexes still remains an open problem.

3 Proposed Method: PPFrame

3.1 Problem Definition

We model a protein as a joint representation of its amino acid sequence and 3D backbone structure. Let $\mathcal{A}$ denote the set of 20 standard amino acids. A protein of length L is represented by a sequence

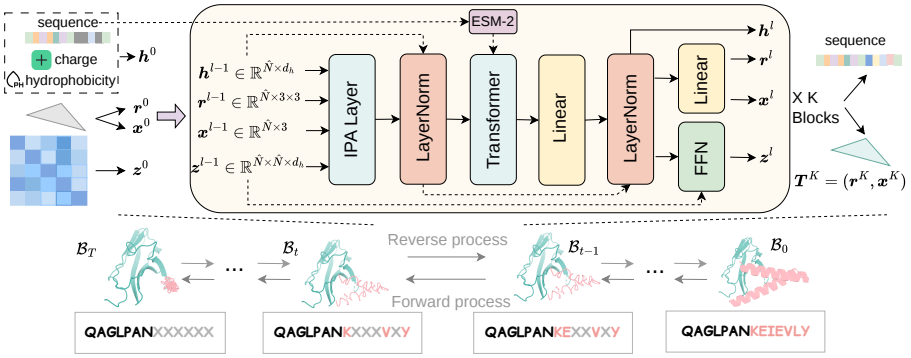


Figure 1: Overview of the proposed PPFrame architecture. At each diffusion time step, the target protein structure is rendered in green, while the binder protein is shown in pink.

$\boldsymbol{a} = \{a_1, a_2, ..., a_L\} \in \mathcal{A}^L$ and an associated backbone structure $\boldsymbol{s}$. Following Jumper et al. (2021), the backbone geometry is parameterized by a collection of local rigid frames mapping from reference coordinates $[N^*, C_\alpha^*, C^*, O^*] \in \mathbb{R}^{4\times 3}$ centered at $C_\alpha^* = (0, 0, 0)$. Specifically, for residue $i \in \{1, \ldots, L\}$, the mainchain heavy atom coordinates $[N_i, C_i, (C_\alpha)_i]$ are represented by applying a rigid body transformation $\boldsymbol{T}_i \in SE(3)$ to the reference frame: $[N_i, C_i, (C_\alpha)_i] = \boldsymbol{T}_i \cdot [N^*, C^*, C_\alpha^*]$. Each frame $\boldsymbol{T}_i = (\boldsymbol{r}_i, \boldsymbol{x}_i)$ comprises a rotation matrix $\boldsymbol{r}_i \in SO(3)$ and a translation vector $\boldsymbol{x}_i \in \mathbb{R}^3$, acting on a spatial coordinate $\boldsymbol{v} \in \mathbb{R}^3$ via $\boldsymbol{T}_i \cdot \boldsymbol{v} = \boldsymbol{r}_i \boldsymbol{v} + \boldsymbol{x}_i$. The full backbone structure is thus defined as $\boldsymbol{s} = \{\boldsymbol{T}_1 \cdot [N^*, C^*, C_\alpha^*], \ldots, \boldsymbol{T}_L \cdot [N^*, C^*, C_\alpha^*]\}$, with backbone oxygen atoms positioned using the torsion angle ψ around the C_α–C bond. Furthermore, each residue i is assigned a physicochemical feature vector $\boldsymbol{f}_i = (y_i, c_i)$ encoding its hydrophobicity y_i and charge c_i, yielding a protein feature sequence $\boldsymbol{f} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_L)$.

Given a target protein $\mathcal{T} = \{\boldsymbol{a}^\mathcal{T}, \boldsymbol{s}^\mathcal{T}, \boldsymbol{f}^\mathcal{T}\}$ of length M, the objective of protein-protein complex co-design is to generate a binder protein $\mathcal{B} = \{\boldsymbol{a}^\mathcal{B}, \boldsymbol{s}^\mathcal{B}\}$ of length N displaying complementary physicochemical and geometric properties to target $\mathcal{T}$. This entails jointly sampling the binder sequence $\boldsymbol{a}^\mathcal{B} = \{a_1^\mathcal{B}, \ldots, a_N^\mathcal{B}\}$ and backbone transformations $\boldsymbol{T}^\mathcal{B} = \{\boldsymbol{T}_1^\mathcal{B}, \ldots, \boldsymbol{T}_N^\mathcal{B}\}$ with aligned physicochemical features $\boldsymbol{f}^\mathcal{B}$. Formally, we formulate this co-design task as learning a conditional generative model $p_\theta(\mathcal{B} \mid \mathcal{T})$ parameterized by θ.

3.2 Overall Model Architecture

The overall architecture of PPFrame is illustrated in Figure 1. PPFrame is a multimodal denoising diffusion probabilistic model (DDPM) (Ho et al., 2020) designed for joint generation of a binder's primary sequence $\boldsymbol{a}^\mathcal{B}$ and backbone structure $\boldsymbol{s}^\mathcal{B}$ conditioned on a target protein $\mathcal{T}$. The framework models a forward perturbation process q that corrupts data state $\mathcal{B}_0 \sim q(\mathcal{B}_0 \mid \mathcal{T})$ into noise $\mathcal{B}_\mathbb{T} \sim q_{\text{noise}}$ over $\mathbb{T}$ steps, alongside a parameterized reverse generative process p_θ:

$$q(\mathcal{B}_{1:\mathbb{T}} \mid \mathcal{B}_0, \mathcal{T}) = \prod_{t=1}^{\mathbb{T}} q(\mathcal{B}_t \mid \mathcal{B}_{t-1}, \mathcal{T}), \quad p_\theta(\mathcal{B}_{0:\mathbb{T}} \mid \mathcal{T}) = p(\mathcal{B}_\mathbb{T}) \prod_{t=1}^{\mathbb{T}} p_\theta(\mathcal{B}_{t-1} \mid \mathcal{B}_t, \mathcal{T}) \tag{1}$$

Each latent variable $\mathcal{B}_t = (\boldsymbol{a}_t^\mathcal{B}, \boldsymbol{s}_t^\mathcal{B})$ represents a hybrid state spanning discrete amino acid categories and continuous $SE(3)$ transformations. Model parameters θ are optimized by maximizing the standard evidence lower bound on log-likelihood:

$$\mathcal{L}_{\text{VLB}} = \mathbb{E}_q \Big[\log p_\theta(\mathcal{B}_0 \mid \mathcal{B}_1, \mathcal{T}) - \sum_{t=2}^{\mathbb{T}} \underbrace{\text{KL}\left[q(\mathcal{B}_{t-1} \mid \mathcal{B}_t, \mathcal{B}_0, \mathcal{T}) \parallel p_\theta(\mathcal{B}_{t-1} \mid \mathcal{B}_t, \mathcal{T})\right]}_{\mathcal{L}_t} \Big] \tag{2}$$

By unifying discrete sequence transitions with continuous manifold diffusion, PPFrame enhances sequence-structure consistency throughout generative sampling.

3.3 PPFrame Diffusion Process

We formulate the joint forward diffusion process of protein residue sequence $\boldsymbol{a}_t^\mathcal{B}$ and backbone structure $\boldsymbol{s}_t^\mathcal{B}$ as factorized transitions across individual residues and modalities at each time step t:

$$q(\mathcal{B}_t \mid \mathcal{B}_{t-1}, \mathcal{T}) = \prod_{i=1}^{N} q(a_{i,t}^\mathcal{B} \mid a_{i,t-1}^\mathcal{B}, \mathcal{T}) \cdot \prod_{i=1}^{N} q(\boldsymbol{s}_{i,t}^\mathcal{B} \mid \boldsymbol{s}_{i,t-1}^\mathcal{B}, \mathcal{T}) \tag{3}$$

where $a^{\mathcal{B}}_{i,t}$ and $\boldsymbol{s}^{\mathcal{B}}_{i,t} = \boldsymbol{T}^{\mathcal{B}}_{i,t} \cdot [N^*, C^*_\alpha, C^*]$ denote the i-th residue type and local rigid frame at time step t, respectively. This factorization enables efficient noise sampling across different modalities.

Discrete Sequence Diffusion. For the sequence components $\boldsymbol{a}^{\mathcal{B}}$, state transitions are defined via transition matrices $[\boldsymbol{Q}_t]_{mn} = q(a^{\mathcal{B}}_{i,t} = n \mid a^{\mathcal{B}}_{i,t-1} = m)$. Expressing $a^{\mathcal{B}}_{i,t}$ as a one-hot row vector $\boldsymbol{a}^{\mathcal{B}}_{i,t}$, the single-step transition is given by:

$$q(\boldsymbol{a}^{\mathcal{B}}_{i,t} \mid \boldsymbol{a}^{\mathcal{B}}_{i,t-1}, \mathcal{T}) = \mathrm{Cat}(\boldsymbol{a}^{\mathcal{B}}_{i,t}; \boldsymbol{p} = \boldsymbol{a}^{\mathcal{B}}_{i,t-1}\boldsymbol{Q}_t) \tag{4}$$

where $\mathrm{Cat}(\boldsymbol{a}; \boldsymbol{p})$ denotes a categorical distribution with probability vector $\boldsymbol{p}$. Starting from $\boldsymbol{a}^{\mathcal{B}}_{i,0}$, the t-step marginal and posterior distributions are analytically derived as:

$$\begin{aligned} q(\boldsymbol{a}^{\mathcal{B}}_{i,t} \mid \boldsymbol{a}^{\mathcal{B}}_{i,0}, \mathcal{T}) &= \mathrm{Cat}(\boldsymbol{a}^{\mathcal{B}}_{i,t}; \boldsymbol{p} = \boldsymbol{a}^{\mathcal{B}}_{i,0}\bar{\boldsymbol{Q}}_t), \quad \text{where} \quad \bar{\boldsymbol{Q}}_t = \textstyle\prod_{j=1}^{t} \boldsymbol{Q}_j \\ q(\boldsymbol{a}^{\mathcal{B}}_{i,t-1} | \boldsymbol{a}^{\mathcal{B}}_{i,t}, \boldsymbol{a}^{\mathcal{B}}_{i,0}, \mathcal{T}) &= \mathrm{Cat}\Big(\boldsymbol{a}^{\mathcal{B}}_{i,t-1}; \boldsymbol{p} = \frac{\boldsymbol{a}^{\mathcal{B}}_{i,t}\boldsymbol{Q}_t^\top \odot \boldsymbol{a}^{\mathcal{B}}_{i,0}\overline{\boldsymbol{Q}}_{t-1}}{\boldsymbol{a}^{\mathcal{B}}_{i,0}\overline{\boldsymbol{Q}}_t \boldsymbol{a}^{\mathcal{B}\top}_{i,t}}\Big) \end{aligned} \tag{5}$$

Following Austin et al. (2021), we adopt an absorbing-state transition matrix $\boldsymbol{Q}_t$ with an absorbing token [mask], where each token either remains unchanged or transitions to [mask] with probability β_t:

$$[\boldsymbol{Q}_t]_{mn} = \begin{cases} 1 & \text{if } m = n = \text{[mask]} \\ 1 - \beta_t & \text{if } m = n \neq \text{[mask]} \\ \beta_t & \text{if } n = \text{[mask]}, m \neq \text{[mask]} \end{cases} \tag{6}$$

Continuous Structure Diffusion on $SE(3)$. To generate backbone structures in $SE(3)^N$, we build upon Riemannian score-based generative modeling (De Bortoli et al., 2022). Following Yim et al. (2023), we treat the rotational $SO(3)$ and translational $\mathbb{R}^3$ sub-manifolds independently, setting the final invariant distribution to $p_{\mathrm{SE}(3)}(\boldsymbol{T}^{\mathcal{B}}_{i,t}) \propto \mathcal{U}_{\mathrm{SO}(3)}(\boldsymbol{r}^{\mathcal{B}}_{i,t})\mathcal{N}(\boldsymbol{x}^{\mathcal{B}}_{i,t}; \boldsymbol{0}, \boldsymbol{I}_3)$. The forward Stochastic Differential Equation (SDE) on $\boldsymbol{T}^{\mathcal{B}}_{i,t} = (\boldsymbol{r}^{\mathcal{B}}_{i,t}, \boldsymbol{x}^{\mathcal{B}}_{i,t})$ is defined by:

$$d\boldsymbol{T}^{\mathcal{B}}_{i,t} = \left[\boldsymbol{0}, -\frac{1}{2}\boldsymbol{x}^{\mathcal{B}}_{i,t}\right] dt + \left[d\boldsymbol{B}^{(i,t)}_{\mathrm{SO}(3)}, d\boldsymbol{B}^{(i,t)}_{\mathbb{R}^3}\right] \tag{7}$$

where $\boldsymbol{B}_{\mathrm{SO}(3)}$ and $\boldsymbol{B}_{\mathbb{R}^3}$ denote Brownian motions on $SO(3)$ and $\mathbb{R}^3$, respectively. The score network $s_\theta(t, \boldsymbol{T}^{\mathcal{B}}_{i,t}; \mathcal{T})$ is trained via denoising score matching:

$$\mathcal{L}_{\text{struct}} = \mathbb{E}\left[\lambda_t \left\| \nabla_{\boldsymbol{T}^{\mathcal{B}}_{i,t}} \log p_{t|0}(\boldsymbol{T}^{\mathcal{B}}_{i,t} \mid \boldsymbol{T}^{\mathcal{B}}_{i,0}; \mathcal{T}) - s_\theta(t, \boldsymbol{T}^{\mathcal{B}}_{i,t}; \mathcal{T}) \right\|^2\right] \tag{8}$$

By leveraging independence across coordinates, the conditional score decomposes into rotational and translational components: $\nabla_{\boldsymbol{T}^{\mathcal{B}}_{i,t}} \log p_{t|0} = [\nabla_{\boldsymbol{r}^{\mathcal{B}}_{i,t}} \log p_{t|0},\ \nabla_{\boldsymbol{x}^{\mathcal{B}}_{i,t}} \log p_{t|0}]$. For $SO(3)$, the perturbation kernel $p_{t|0}$ is governed by the heat kernel on $SO(3)$ (Hsu, 2002), enabling analytical score evaluation:

$$\begin{aligned} \nabla_{\boldsymbol{r}^{\mathcal{B}}_{i,t}} \log p_{t|0}(\boldsymbol{r}^{\mathcal{B}}_{i,t} \mid \boldsymbol{r}^{\mathcal{B}}_{i,0}, \mathcal{T}) &= \frac{\boldsymbol{r}^{\mathcal{B}}_{i,t}}{\omega_{i,t}} \log\left((\boldsymbol{r}^{\mathcal{B}}_{i,0})^\top \boldsymbol{r}^{\mathcal{B}}_{i,t}\right) \frac{\partial_\omega f(\omega_{i,t}, t)}{f(\omega_{i,t}, t)} \\ f(\omega, t) &= \sum_{l=0}^{\infty} (2l+1) e^{-l(l+1)t/2} \frac{\sin((l+1/2)\omega)}{\sin(\omega/2)} \end{aligned} \tag{9}$$

where $\omega_{i,t} = \omega((\boldsymbol{r}^{\mathcal{B}}_{i,0})^\top \boldsymbol{r}^{\mathcal{B}}_{i,t})$ is the rotation angle in radians. For $\mathbb{R}^3$, the translation process follows an Ornstein-Uhlenbeck process (Song et al., 2020) with kernel $p_{t|0}(\boldsymbol{x}^{\mathcal{B}}_{i,t} \mid \boldsymbol{x}^{\mathcal{B}}_{i,0}) = \mathcal{N}(\boldsymbol{x}^{\mathcal{B}}_{i,t}; e^{-t/2}\boldsymbol{x}^{\mathcal{B}}_{i,0}, (1 - e^{-t})\boldsymbol{I}_3)$, yielding the exact conditional score:

$$\nabla_{\boldsymbol{x}^{\mathcal{B}}_{i,t}} \log p_{t|0}(\boldsymbol{x}^{\mathcal{B}}_{i,t} \mid \boldsymbol{x}^{\mathcal{B}}_{i,0}, \mathcal{T}) = (1 - e^{-t})^{-1}\left(e^{-t/2}\boldsymbol{x}^{\mathcal{B}}_{i,0} - \boldsymbol{x}^{\mathcal{B}}_{i,t}\right) \tag{10}$$

3.4 IPAFormer Architecture for the Reverse Generative Process

The reverse generative process parameterizes the conditional transition distribution $p_\theta(\mathcal{B}_{t-1} \mid \mathcal{B}_t, \mathcal{T})$ to reconstruct uncorrupted sequence-structure pairs from noisy states. We parameterize this reverse mapping using IPAFormer, a multi-block neural architecture designed for joint sequence-structure co-design. IPAFormer iteratively refines spatial frames and sequence representations across K stacked

blocks. Each block interleaves spatial-based invariant point attention (IPA) layer (Jumper et al., 2021) with sequence-based self-attention layers, effectively modeling 3D geometric interactions alongside global amino acid dependencies.

Node embeddings $\boldsymbol{h}^0 = (\boldsymbol{h}_1^0, \dots, \boldsymbol{h}_{\hat{N}}^0) \in \mathbb{R}^{\hat{N} \times d_h}$ are initialized with residue-level physicochemical features, while edge embeddings $\boldsymbol{z}^0 \in \mathbb{R}^{\hat{N} \times \hat{N} \times d_z}$ capture pairwise initial geometry following Trippe et al. (2022), where $\hat{N} = M + N$ denotes the total residue length of the target-binder complex. To enhance sequence modeling capacity, we incorporate pretrained residue representations $\boldsymbol{e} \in \mathbb{R}^{\hat{N} \times d_e}$ extracted from ESM-2 (Lin et al., 2022). Within the l-th block ($l \in \{1, \dots, K\}$), representations and rigid frames are updated as follows:

$$\begin{aligned}
\boldsymbol{h}^{(l-0.75)} &= \text{IPALayer}\big(\boldsymbol{h}^{(l-1)}, \boldsymbol{z}^{(l-1)}, \boldsymbol{T}^{(l-1)}\big), \quad \boldsymbol{h}^{(l-0.5)} = \text{LayerNorm}\big(\boldsymbol{h}^{(l-0.75)} + \boldsymbol{h}^{(l-1)}\big) \\
\boldsymbol{h}^{(l-0.25)} &= \text{Transformer}\big(\boldsymbol{h}^{(l-0.5)} + \boldsymbol{e}\big), \quad \boldsymbol{h}^{l} = \text{LayerNorm}\big(\text{Linear}(\boldsymbol{h}^{(l-0.25)}) + \boldsymbol{h}^{(l-0.5)}\big) \\
\tilde{\boldsymbol{T}}^l &= \boldsymbol{h}^l \boldsymbol{W}_f, \quad \boldsymbol{r}^l = \boldsymbol{r}^{l-1} + \boldsymbol{r}^{l-1} \cdot \tilde{\boldsymbol{T}}^l_{[:3]}, \quad \boldsymbol{x}^l = \boldsymbol{x}^{l-1} + \boldsymbol{r}^{l-1} \cdot \tilde{\boldsymbol{T}}^l_{[3:]}, \quad \boldsymbol{T}^l = (\boldsymbol{r}^l, \boldsymbol{x}^l) \\
\boldsymbol{z}^l &= \text{FFN}\big(f(\boldsymbol{h}^l, \boldsymbol{z}^{(l-1)})\big), \quad f_{ij}(\boldsymbol{h}^l, \boldsymbol{z}^{(l-1)}) = \text{Concatenate}\big(\text{Linear}(\boldsymbol{h}_i^l), \text{Linear}(\boldsymbol{h}_j^l), \boldsymbol{z}_{ij}^{(l-1)}\big)
\end{aligned} \tag{11}$$

where $\boldsymbol{W}_f \in \mathbb{R}^{d_h \times 6}$ is a trainable projection matrix, and $\boldsymbol{T}^0 = [\boldsymbol{T}^{\mathcal{T}}, \tilde{\boldsymbol{T}}^{\mathcal{B}}]$ concatenates ground-truth target frames $\boldsymbol{T}^{\mathcal{T}}$ with perturbed binder frames $\tilde{\boldsymbol{T}}^{\mathcal{B}}$. FFN denotes the feed-forward layer.

From the final outputs $(\boldsymbol{h}^K, \boldsymbol{T}^K)$, we compute rotational and translational scores, predict the uncorrupted sequence distribution, and estimate backbone ψ torsion angles:

$$s_\theta^{r,i} = f_r(\tilde{\boldsymbol{r}}_i^{\mathcal{B}}, \boldsymbol{r}_i^{\mathcal{B},K}), \quad s_\theta^{x,i} = f_x(\tilde{\boldsymbol{x}}_i^{\mathcal{B}}, \boldsymbol{x}_i^{\mathcal{B},K}), \quad p_\theta(\hat{\boldsymbol{a}}_{i,0}^{\mathcal{B}}) = \text{Softmax}\big(\text{Linear}(\boldsymbol{h}_i^K)\big), \quad \psi_i = f_\psi(\boldsymbol{h}_i^K) \tag{12}$$

where f_r and f_x compute score function approximations from noisy frames $(\tilde{\boldsymbol{r}}_i^{\mathcal{B}}, \tilde{\boldsymbol{x}}_i^{\mathcal{B}})$ and network-predicted frames $(\boldsymbol{r}_i^{\mathcal{B},K}, \boldsymbol{x}_i^{\mathcal{B},K})$, and f_ψ maps node representations to ψ torsion angles. The predicted distribution $p_\theta(\hat{\boldsymbol{a}}_{i,0}^{\mathcal{B}})$ parameterizes the clean sequence $\boldsymbol{a}_{i,0}^{\mathcal{B}}$, allowing analytical calculation of the reverse categorical posterior $p_\theta(\boldsymbol{a}_{i,t-1}^{\mathcal{B}} | \boldsymbol{a}_{i,t}^{\mathcal{B}}, \hat{\boldsymbol{a}}_{i,0}^{\mathcal{B}}, \mathcal{T})$.

3.5 Training Objective

To achieve effective sequence-structure co-design, PPFrame is optimized via a multitask learning objective $\mathcal{L}_t$ that combines discrete sequence learning, continuous $SE(3)$ score-matching frame losses, and late-stage auxiliary physical regularization with weights α, γ, and ω:

$$\begin{aligned}
\mathcal{L}_t &= \alpha \mathcal{L}_{\text{seq}} + \gamma(\mathcal{L}_{\text{struct}}^r + \mathcal{L}_{\text{struct}}^x) + \omega \mathcal{L}_{\text{aux}} \\
\mathcal{L}_{\text{seq}} &= \sum\nolimits_{i=1}^N \text{KL}\left(q(\boldsymbol{a}_{i,t-1}^{\mathcal{B}} \mid \boldsymbol{a}_{i,t}^{\mathcal{B}}, \boldsymbol{a}_{i,0}^{\mathcal{B}}, \mathcal{T}) \;\|\; p_\theta(\boldsymbol{a}_{i,t-1}^{\mathcal{B}} | \boldsymbol{a}_{i,t}^{\mathcal{B}}, \hat{\boldsymbol{a}}_{i,0}^{\mathcal{B}}, \mathcal{T})\right) \\
\mathcal{L}_{\text{struct}}^r &= \sum\nolimits_{i=1}^N \left\| \nabla_{\boldsymbol{r}_{i,t}^{\mathcal{B}}} \log p_{t|0}(\boldsymbol{r}_{i,t}^{\mathcal{B}} \mid \boldsymbol{r}_{i,0}^{\mathcal{B}}, \mathcal{T}) - s_\theta^{r,i} \right\|^2, \quad \mathcal{L}_{\text{struct}}^x = \sum\nolimits_{i=1}^N \left\| \boldsymbol{x}_i^{\mathcal{B}} - \boldsymbol{x}_i^{\mathcal{B},K} \right\|^2
\end{aligned} \tag{13}$$

where $\mathcal{L}_{\text{seq}}$ maximizes the variational lower bound in Eq. 2, and the frame losses minimize the score-matching objective in Eq. 8. By applying target weighting schedules $\lambda_{i,t}^r = 1/\mathbb{E}\big[\|\nabla \log p_{t|0}(\boldsymbol{r}_{i,t} \mid \boldsymbol{r}_{i,0})\|_{\text{SO}(3)}^2\big]$ for rotations and $\lambda_{i,t}^x = (1 - e^{-t})/e^{-t/2}$ for translations, the frame objectives simplify to the unweighted formulations above.

To enforce physical contact fidelity during final structure generation, we activate two auxiliary terms, $\mathcal{L}_{\text{aux}} = \mathcal{L}_{\text{struct}}^{\text{atom}} + \mathcal{L}_{\text{struct}}^{\text{pair}}$, during late denoising time steps ($t < 0.25$):

$$\begin{aligned}
\mathcal{L}_{\text{struct}}^{\text{atom}} &= \frac{1}{4N} \sum_{i=1}^N \sum_{m \in \mathbb{A}} \left\| \boldsymbol{x}_{i,m}^{\mathcal{B}} - \boldsymbol{x}_{i,m}^{\mathcal{B},K} \right\|^2, \quad \mathbb{A} = \{\text{N}, \text{C}_\alpha, \text{C}, \text{O}\} \\
\mathcal{L}_{\text{struct}}^{\text{pair}} &= \frac{1}{S} \sum_{1 \le i \ne j \le N} \sum_{m,n \in \mathbb{A}} \mathbb{I}(d_{mn}^{ij} < \tau) \left\| d_{mn}^{ij} - \hat{d}_{mn}^{ij} \right\|^2, \quad S = \sum_{1 \le i \ne j \le N} \sum_{m,n \in \mathbb{A}} \mathbb{I}(d_{mn}^{ij} < \tau)
\end{aligned} \tag{14}$$

where d^* denotes the pairwise atomic distances, and τ specifies the interaction distance cutoff.

3.6 Self-Conditioned Inference

During inference, conditioned on the target protein's sequence $\boldsymbol{a}^{\mathcal{T}}$, structure $\boldsymbol{s}^{\mathcal{T}}$, and physicochemical features $\boldsymbol{f}^{\mathcal{T}}$, PPFrame conditionally co-samples the binder sequence $\boldsymbol{a}^{\mathcal{B}}$ and backbone structure

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.523	81.61	21.73%	37.90%	73.01%	-45.97	50.00%	50.00%
APM	0.252	77.66	2.17%	31.26%	53.27%	-45.12	23.91%	34.78%
Complexa	0.521	84.61	30.43%	**53.77%**	**82.69%**	-56.57	56.52%	54.34%
PPFrame	**0.592**	**87.18**	**47.83%**	33.22%	57.49%	**-67.80**	**65.22%**	**65.22%**

Table 1: Model performance on general protein-protein complex design task. Our PPFrame achieves the best sequence design success rate, structure binding energy and sequence-structure co-designability.

$\boldsymbol{s}^{\mathcal{B}}$ to yield high-affinity complexes. Sequence generation initializes all binder positions in an absorbing state [mask], fixing residue types and updating physicochemical features $\boldsymbol{f}^{\mathcal{B}}$ upon state transition. Structure generation utilizes Euler-Maruyama geodesic random walks on $SE(3)$ over $\mathbb{T}$ steps (De Bortoli et al., 2022). Translations receive Gaussian updates from $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_3)$, while rotations update along the Lie algebra $SO(3)$ tangent space as $\boldsymbol{r}'_{i,t} = \boldsymbol{r}_{i,t} \sum_{j=1}^{3} \delta_j \boldsymbol{e}_j$ with $\delta_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ are orthonormal basis vectors. To maximize structural fidelity, we implement a self-conditioned inference scheme at each time step t: IPAFormer first performs an intermediate forward pass to update C_α positions, then recalculates spatial pair embeddings to refine sequence likelihoods and rigid frames in a second pass.

4 Experiments

4.1 PPIBench Construction

To establish a foundation model for protein-protein complex design that can be seamlessly adapted to real-world applications, we curate a comprehensive dataset called PPIBench. Following the data processing pipeline of AlphaFold3 (Abramson et al., 2024), we extract 301,182 chain-pair interfaces from the PDB. Complete processing workflows are detailed in Appendix A.1. By treating both chains within each interface symmetrically as target-binder pairs, we construct a benchmark comprising 602,364 complexes across diverse scaffolds. To prevent data leakage, sequences are clustered at 50% identity, yielding 28,884 distinct clusters. We reserve 10 clusters each for validation and testing, allocating the remaining clusters for training. Detailed data statistics are given in Appendix Table 4.

4.2 Experimental Setup

Implementation Details. The IPAFormer comprises five blocks, each containing one IPA layer and six self-attention layers. Parameters for the self-attention layers are initialized from the pretrained 150M ESM-2 model. Node and edge embedding dimensions are set to 640 and 320, respectively, bringing the total parameter count of PPFrame to 240M. The diffusion process is configured for 1,000 steps. In inference, for each target protein, we generate 10 binder candidates and rank them by their complex pLDDT scores, selecting the top-k candidates. More implementation details are provided in Appendix B.

Baseline Models. We compare the proposed PPFrame against the following representative protein complex co-design methods: (1) **RFD+ProteinMPNN (IF)** first applies RFDiffusion (Watson et al., 2023) to design a binder protein structure based on the given target and then uses ProteinMPNN (Dauparas et al., 2022) to generate the corresponding sequence. (2) **APM** (Chen et al., 2025) is an all-atom generative model for protein complex design. (3) **Proteína-Complexa** (Didi et al., 2026) is a fully atomistic binder generation method showing strong performance on protein-binding protein design.

Evaluation Metrics. To comprehensively evaluate PPFrame, we assess generated complexes across three dimensions: sequence quality, structure reliability, and sequence-structure consistency. **Sequence Evaluation:** We evaluate designed sequences using five metrics: (1) **ipTM** via AlphaFold3 (AF3) to measure interface confidence; (2) **pLDDT** via AF3 to assess complex structure stability; (3) **Success Rate (SR)** defined as the fraction of designed candidates that simultaneously achieve an $\mathrm{ipTM} > 0.8$ and a $\mathrm{pLDDT} > 80$; (4) **Diversity**, measured as the amino acid variation among candidates for the same target; and (5) **Novelty**, calculated as the sequence dissimilarity between candidates and reference sequences. **Structure Evaluation:** We assess designed structures via:

Models	FGFR2 (Seen)			InsulinR (Seen)			PDGFR (Seen)		
	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)
RFD+IF	36.84%	-42.04	63.15%	56.52%	-35.35	65.21%	3.84%	-59.37	53.84%
APM	29.82%	-37.02	54.38%	43.47%	-31.04	82.61%	0.0	-61.54	61.53%
Complexa	54.38%	-45.81	73.68%	78.26%	-41.32	86.95%	**19.23%**	-66.56	**88.46%**
PPFrame	**61.40%**	**-50.87**	**98.24%**	**86.95%**	**-46.92**	**100%**	7.69%	**-68.82**	30.76%
Models	TGFb (Seen)			VirB8 (Seen)			EGFR (Zero-shot)		
	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)
RFD+IF	11.11%	-31.09	**55.56%**	14.28%	-32.91	57.14%	**25.00%**	-46.87	25.00%
APM	11.11%	-23.16	11.11%	0.0	-19.59	42.85%	0.0	-32.76	50.00%
Complexa	22.22%	-46.54	33.33%	14.28%	-38.56	85.71%	0.0	-40.90	50.00%
PPFrame	**22.22%**	**-49.29**	22.22%	**14.28%**	**-48.31**	**100%**	0.0	**-48.35**	**100.0%**
Models	H3 (Zero-shot)			IL7Ra (Zero-shot)			Tie2 (Zero-shot)		
	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)
RFD+IF	52.63%	-35.93	26.31%	14.28%	-40.08	42.85%	**50.00%**	-38.95	75.00%
APM	47.38%	-36.45	42.10%	0.0	-23.93	57.14%	0.0	-27.85	50.00%
Complexa	60.52%	-40.65	**47.36%**	28.57%	-42.26	71.42%	0.0	-33.83	50.00%
PPFrame	**97.36%**	**-47.08**	31.57%	**42.85%**	**-50.42**	**100%**	0.0	**-42.39**	**100.00%**
Models	TrkA (Zero-shot)			Average			Novelty (↑)	Diversity (↑)	
	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)			
RFD+IF	25.00%	-33.28	50.00%	28.95%	-39.59	51.41%	70.69%	30.38%	
APM	25.00%	-31.92	25.00%	15.67%	-32.53	47.67%	51.47%	29.02%	
Complexa	50.00%	-41.50	50.00%	32.74%	-43.79	63.69%	78.69%	**42.69%**	
PPFrame	**75.00%**	**-46.04**	**75.00%**	**40.77%**	**-49.85**	**75.78%**	**81.24%**	29.08%	

Table 2: Model performance on target-protein mini-binder complex design task. Consistency denotes sequence-structure co-designability. Our PPFrame achieves the best average sequence design success rate, structure binding energy and sequence-structure co-designability.

(6) **Binding Energy** (ΔG), computed via Rosetta (Chaudhury et al., 2010) to estimate interface energetic stability, where lower values indicate stronger binding propensity, and (7) **Designability**, defined as the fraction of designs achieving a self-consistency RMSD (scRMSD) < 2 Å (Trippe et al., 2022). **Sequence-Structure Consistency:** We measure joint consistency via (8) **Co-designability (Consistency)**, defined as the proportion of generated binder sequences whose AF3-folded structure matches the corresponding designed structures within 2 Å RMSD (Didi et al., 2026).

4.3 General Protein-Protein Complex Design

This task evaluates the foundational capability of PPFrame to design protein-protein complexes across diverse target scaffolds. We train and evaluate PPFrame on our curated PPIBench. Top-1 candidate performance for all methods is summarized in Table 1. **PPFrame demonstrates the best performance in generating high-affinity, structurally consistent binders across a wide spectrum of targets.** Specifically, PPFrame achieves top performance across all metrics spanning sequence quality, structural reliability, and sequence-structure co-designability, reaching a sequence design success rate of 47.83%, an interface binding energy of -67.8, and a co-designability score of 65.22%. Although sequence diversity and novelty scores are slightly lower than those of Proteína-Complexa, PPFrame achieves higher structural diversity (Appendix Table 7). This indicates that PPFrame successfully generates binders with flexible loop configurations and distinct subdomain topologies, whereas binders from Proteína-Complexa have almost invariant secondary structure cores.

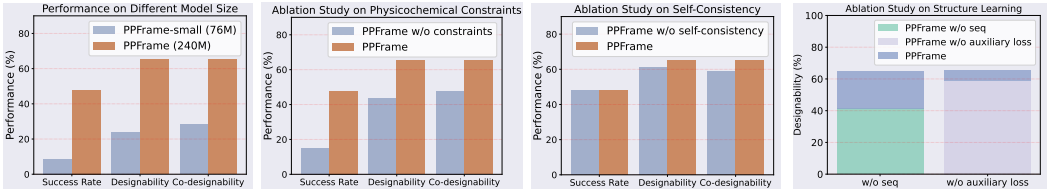
4.4 Target Protein-Mini Binder Complex Design

Datasets. We curate experimentally validated target-protein mini-binder complexes across ten diverse target classes from Bennett et al. (2023). For targets with more than 50 complexes, we partition the data using a random 8:1:1 train/validation/test split. For targets with fewer complexes, all instances are reserved for testing to evaluate model generalization in a zero-shot setting. Comprehensive dataset statistics are provided in Appendix Table 5.

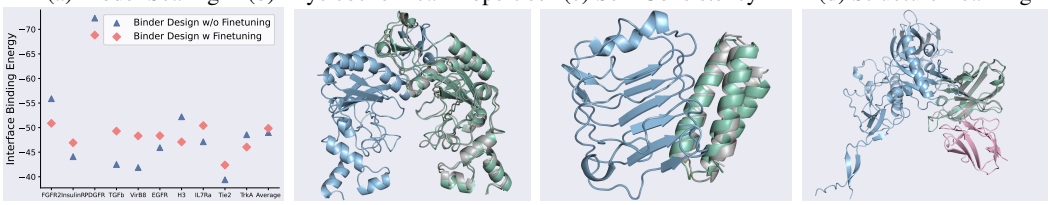
Main Results. The results for target-protein mini-binder complex design are summarized in Table 2, with target-specific breakdowns provided in Appendix Tables 8–17. **Overall, PPFrame demon-**

Models	Sequence						Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	H1-Novelty (↑)	H2-Novelty (↑)	H3-Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.508	80.05	16.79%	47.30%	56.22%	61.98%	-33.09	59.29%	59.29%
APM	0.256	77.92	6.61%	45.71%	54.17%	64.43%	-26.68	53.57%	52.74%
Complexa	0.515	81.08	16.92%	**53.80%**	66.26%	71.91%	-32.80	90.26%	92.32%
PPFrame	**0.526**	**81.41**	**18.77%**	52.31%	**67.85%**	**71.93%**	**-44.46**	**98.34%**	**99.39%**

Table 3: Model performance on antigen-antibody complex design task. Our PPFrame achieves the best sequence design success rate, structure binding energy and sequence-structure co-designability.



(a) Model Scaling (b) Physicochemical Properties (c) Self-Consistency (d) Structure Learning



(e) Binder Finetuning (f) General Protein Complex (g) Binder Design (h) Antibody Design

Figure 2: Ablation study: (a) - (e), and case study with target protein (f) PDB 1STX.E, (g) InsulinR, (h) HIV-1 (7UCG). Target proteins are shown in blue, the folded structures for the designed binder sequences in green, and the designed binder structures in gray, with light chain in pink for antibody.

strates superior capacity for generating novel, high-affinity binders across diverse target classes. Specifically, PPFrame achieves higher sequence design success rates on 7 out of 10 targets, superior structure binding energy across all 10 targets, and improved co-designability on 7 out of 10 targets. Notably, PPFrame also yields higher novelty scores, reversing the trend observed in our general complex design task. We attribute this shift to the distinctions between the fine-tuning dataset and PPIBench, enabling PPFrame to navigate a broader candidate space for novel binder discovery.

4.5 Antigen-Antibody Complex Design

Dataset. Following Jin et al. (2022), we curate 4,261 complete antigen-antibody complexes from the Structural Antibody Database (Dunbar et al. (2014), SAbDab) by removing cases lacking a light chain or antigen. Train/validation/test splits (8:1:1 ratio) are constructed by clustering individual CDR sequences at 40% identity using MMseqs2 (Steinegger & Söding, 2017), yielding 1,034 (CDR-H1), 1,418 (CDR-H2), and 2,254 (CDR-H3) clusters. Full split statistics are provided in Appendix Table 6.

Main Results. Average performance across distinct CDR cluster splits for antigen-antibody complex design is summarized in Table 3, with CDR-specific breakdowns detailed in Appendix Tables 18–20. **Overall, PPFrame outperforms all baselines in generating high-affinity and novel antibodies.** Specifically, PPFrame achieves top performance across sequence design success rate, CDR H2/H3 novelty, interface binding energy (ΔG), and sequence-structure co-designability. Crucially, both the designability and co-designability of PPFrame exceed 95%, demonstrating exceptional structure fidelity and tight sequence-structure consistency.

5 Analysis

5.1 Does Model Scale Help?

PPFrame scales effectively with parameter capacity. We benchmark our 240M PPFrame against a smaller model (76M) configured with 4 blocks and a hidden dimension of 480. As shown in Fig-

ure 2(a), scaling capacity improves sequence success rate, structural designability, and co-designability on PPIBench. These results highlight the scalability of our architecture, indicating that performance scales predictably with model capacity.

5.2 Effect of Physicochemical Complementarity Constraints

As stable protein-protein interactions depend on complementary geometries and physicochemical properties, we conduct an ablation study by training a variant of PPFrame without physicochemical features on PPIBench. As illustrated in Figure 2(b), omitting these constraints leads to a significant performance degradation across both sequence and structural metrics. **These findings confirm that explicitly incorporating physicochemical features is critical for stable interface design.**

5.3 Influence of Self-Consistency

To evaluate the role of self-consistency during inference, we conduct an ablation by disabling this mechanism on the general protein-protein complex design task. As shown in Figure 2(c), removing self-consistency yields a measurable drop in both structural designability and co-designability. **This validates self-consistency as an effective inference-time strategy for generating reliable structures and reinforcing sequence-structure co-design alignment.**

5.4 Impact of Joint Co-Design and Auxiliary Losses

To isolate the contributions of the joint sequence-structure co-design mechanism and auxiliary supervision, we evaluate a structure-only ablation (PPFrame w/o seq) alongside a model variant trained without auxiliary loss terms (PPFrame w/o auxiliary loss). As illustrated in Figure 2(d), both components enhance structural designability, with sequence-structure co-design exerting a significantly stronger effect. We attribute these gains to the stricter physical regularization provided by both mechanisms, **confirming that they are vital for designing reliable structures.**

5.5 Impact of Supervised Fine-Tuning

To evaluate how domain-specific adaptation affects downstream performance, we compare our pretrained PPFrame on the mini-binder design task. As shown in Figure 2(e), targeted fine-tuning improves interface binding energy (ΔG) across 6 out of 10 targets, slightly promoting the mean energy score from -48.98 to -49.85. Full results are provided in Appendix Table 21. **These results demonstrate that while our pretrained model already possesses robust zero-shot design capabilities, supervised fine-tuning further enhances binding affinity for specialized applications.**

5.6 Designing Diverse and High-Affinity Protein Complexes

Figures 2(f)–(h) showcase qualitative case studies across general complex, mini-binder, and antibody design tasks. Spanning distinct target topologies, the generated candidates consistently satisfy stringent quality thresholds: ipTM > 0.8, pLDDT > 85, interface binding energy < -35, and co-design RMSD < 2 Å. **This highlights PPFrame's ability to co-design highly consistent binder sequences and structures against diverse target scaffolds.**

6 Conclusion

In this work, we introduce PPFrame, a physicochemically enhanced multimodal diffusion method for joint binder protein sequence-structure co-design. Powered by IPAFormer, an architecture combining IPA layer with sequence-level self-attention, PPFrame effectively couples 3D geometry with amino acid dependencies. Across general protein complex generation as well as specialized mini-binder and antibody design tasks, PPFrame consistently surpasses state-of-the-art baselines across multiple metrics. Overall, PPFrame demonstrates strong potential as a robust and scalable tool for computational binder discovery. A potential limitation is the absence of wet-lab validation, which is an essential step planned for our future research to further validate the model's practical utility.

AI USE STATEMENT

In this work, we use generative AI tools for polishing writing. Other than that, all work are done by ourselves. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

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*, pp. 1–3, 2024.

Sarah Alamdari, Nitya Thakkar, Rianne Van Den Berg, Neil Tenenholtz, Robert Strome, Alan M Moses, Alex X Lu, Nicolò Fusi, Ava P Amini, and Kevin K Yang. Protein generation with evolutionary diffusion: sequence is all you need. *BioRxiv*, pp. 2023–09, 2023.

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in neural information processing systems*, 34:17981–17993, 2021.

Nathaniel R Bennett, Brian Coventry, Inna Goreshnik, Buwei Huang, Aza Allen, Dionne Vafeados, Ying Po Peng, Justas Dauparas, Minkyung Baek, Lance Stewart, et al. Improving de novo protein binder design with deep learning. *Nature Communications*, 14(1):2625, 2023.

Stephanie Berger, Franziska Seeger, Ta-Yi Yu, Merve Aydin, Huilin Yang, Daniel Rosenblum, Laure Guenin-Macé, Caleb Glassman, Lauren Arguinchona, Catherine Sniezek, et al. Preclinical proof of principle for orally delivered th17 antagonist miniproteins. *Cell*, 187(16):4305–4317, 2024.

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.

Donal J Brennan, Darran P O'connor, Elton Rexhepaj, Fredrik Ponten, and William M Gallagher. Antibody-based proteomics: fast-tracking molecular diagnostics in oncology. *Nature Reviews Cancer*, 10(9):605–617, 2010.

Patrick Bryant, Gabriele Pozzati, and Arne Elofsson. Improved prediction of protein-protein interactions using alphafold2. *Nature communications*, 13(1):1265, 2022.

Longxing Cao, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Joon Sung Park, Kevin M Jude, Iva Marković, Rameshwar U Kadam, Koen HG Verschueren, et al. Design of protein-binding proteins from the target structure alone. *Nature*, 605(7910):551–560, 2022.

Ginger Chao, Wai L Lau, Benjamin J Hackel, Stephen L Sazinsky, Shaun M Lippow, and K Dane Wittrup. Isolating and engineering human antibodies using yeast surface display. *Nature protocols*, 1(2):755–768, 2006.

Sidhartha Chaudhury, Sergey Lyskov, and Jeffrey J Gray. Pyrosetta: a script-based interface for implementing molecular modeling algorithms using rosetta. *Bioinformatics*, 26(5):689–691, 2010.

Ruizhe Chen, Dongyu Xue, Xiangxin Zhou, Zaixiang Zheng, Xiangxiang Zeng, and Quanquan Gu. An all-atom generative model for designing protein complexes. *arXiv preprint arXiv:2504.13075*, 2025.

Aaron Chevalier, Daniel-Adriano Silva, Gabriel J Rocklin, Derrick R Hicks, Renan Vergara, Patience Murapa, Steffen M Bernard, Lu Zhang, Kwok-Ho Lam, Guorui Yao, et al. Massively parallel de novo protein design for targeted therapeutics. *Nature*, 550(7674):74–79, 2017.

Cyrus Chothia and Joël Janin. Principles of protein–protein recognition. *Nature*, 256(5520):705–708, 1975.

Loredana Lo Conte, Cyrus Chothia, and Joël Janin. The atomic structure of protein-protein recognition sites. *Journal of molecular biology*, 285(5):2177–2198, 1999.

Bruno E Correia, John T Bates, Rebecca J Loomis, Gretchen Baneyx, Chris Carrico, Joseph G Jardine, Peter Rupert, Colin Correnti, Oleksandr Kalyuzhniy, Vinayak Vittal, et al. Proof of principle for epitope-focused vaccine design. *Nature*, 507(7491):201–206, 2014.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.

Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. *Advances in neural information processing systems*, 35:2406–2422, 2022.

Kieran Didi, Zuobai Zhang, Guoqing Zhou, Danny Reidenbach, Zhonglin Cao, Sooyoung Cha, Tomas Geffner, Christian Dallago, Jian Tang, Michael Bronstein, et al. Scaling atomistic protein binder design with generative pretraining and test-time compute. In *International Conference on Learning Representations*, volume 2026, pp. 61056–61109, 2026.

James Dunbar, Konrad Krawczyk, Jinwoo Leem, Terry Baker, Angelika Fuchs, Guy Georges, Jiye Shi, and Charlotte M Deane. Sabdab: the structural antibody database. *Nucleic acids research*, 42 (D1):D1140–D1146, 2014.

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*, pp. 2021–10, 2021.

Sarel J Fleishman, Timothy A Whitehead, Damian C Ekiert, Cyrille Dreyfus, Jacob E Corn, Eva-Maria Strauch, Ian A Wilson, and David Baker. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*, 332(6031):816–821, 2011.

Pablo Gainza, Sarah Wehrle, Alexandra Van Hall-Beauvais, Anthony Marchand, Andreas Scheck, Zander Harteveld, Stephen Buckley, Dongchun Ni, Shuguang Tan, Freyr Sverrisson, et al. De novo design of protein interactions with learned surface fingerprints. *Nature*, 617(7959):176–184, 2023.

Benjamin J Hackel, Atul Kapila, and K Dane Wittrup. Picomolar affinity fibronectin domains engineered utilizing loop length diversity, recursive mutagenesis, and loop shuffling. *Journal of molecular biology*, 381(5):1238–1252, 2008.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Elton P Hsu. *Stochastic analysis on manifolds*. Number 38. American Mathematical Soc., 2002.

Ian R Humphreys, Jimin Pei, Minkyung Baek, Aditya Krishnakumar, Ivan Anishchenko, Sergey Ovchinnikov, Jing Zhang, Travis J Ness, Sudeep Banjade, Saket R Bagde, et al. Computed structures of core eukaryotic protein complexes. *Science*, 374(6573):eabm4805, 2021.

Wengong Jin, Jeremy Wohlwend, Regina Barzilay, and Tommi S Jaakkola. Iterative refinement graph neural network for antibody sequence-structure co-design. In *International Conference on Learning Representations*, 2022.

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.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Sal Candido, et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *bioRxiv*, 2022.

Aaron L Nelson, Eugen Dhimolea, and Janice M Reichert. Development trends for human monoclonal antibody therapeutics. *Nature reviews drug discovery*, 9(10):767–774, 2010.

Martin Pacesa, Lennart Nickel, Christian Schellhaas, Joseph Schmidt, Ekaterina Pyatova, Lucas Kissling, Patrick Barendse, Jagrity Choudhury, Srajan Kapoor, Ana Alcaraz-Serna, et al. Bindcraft: one-shot design of functional protein binders. *bioRxiv*, pp. 2024–09, 2024.

Isaac Sappington, Martin Toul, David S Lee, Stephanie A Robinson, Inna Goreshnik, Clara McCurdy, Tung Ching Chan, Nic Buchholz, Buwei Huang, Dionne Vafeados, et al. Improved protein binder design using β-pairing targeted rfdiffusion. *Nature Communications*, 17(1):1101, 2026.

Fabian Sesterhenn, Che Yang, Jaume Bonet, Johannes T Cramer, Xiaolin Wen, Yimeng Wang, Chi-I Chiang, Luciano A Abriata, Iga Kucharska, Giacomo Castoro, et al. De novo protein design enables the precise induction of rsv-neutralizing antibodies. *Science*, 368(6492):eaay5051, 2020.

Daniel-Adriano Silva, Shawn Yu, Umut Y Ulge, Jamie B Spangler, Kevin M Jude, Carlos Labão-Almeida, Lestat R Ali, Alfredo Quijano-Rubio, Mikel Ruterbusch, Isabel Leung, et al. De novo design of potent and selective mimics of il-2 and il-15. *Nature*, 565(7738):186–191, 2019.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. PMLR, 2015.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

Zhenqiao Song, Tiaoxiao Li, Lei Li, and Martin Renqiang Min. Ppdiff: Diffusing in hybrid sequence-structure space for protein-protein complex design. *arXiv preprint arXiv:2506.11420*, 2025.

Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy O'Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, et al. 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.

Brian L Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*, 2022.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Rodrigo Vazquez-Lombardi, Tri Giang Phan, Carsten Zimmermann, David Lowe, Lutz Jermutus, and Daniel Christ. Challenges and opportunities for non-antibody scaffold drugs. *Drug discovery today*, 20(10):1271–1283, 2015.

Susana Vázquez Torres, Philip JY Leung, Preetham Venkatesh, Isaac D Lutz, Fabian Hink, Huu-Hien Huynh, Jessica Becker, Andy Hsien-Wei Yeh, David Juergens, Nathaniel R Bennett, et al. De novo design of high-affinity binders of bioactive helical peptides. *Nature*, 626(7998):435–442, 2024.

Xinyou Wang, Zaixiang Zheng, Fei Ye, Dongyu Xue, Shujian Huang, and Quanquan Gu. Diffusion language models are versatile protein learners. *arXiv preprint arXiv:2402.18567*, 2024.

Xinyou Wang, Zaixiang Zheng, Fei Ye, Dongyu Xue, Shujian Huang, and Quanquan Gu. Dplm-2: A multimodal diffusion protein language model. In *International Conference on Learning Representations*, volume 2025, pp. 35463–35490, 2025.

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.

Kejia Wu, Hanlun Jiang, Derrick R Hicks, Caixuan Liu, Edin Muratspahić, Theresa A Ramelot, Yuexuan Liu, Kerrie McNally, Amit Gaur, Brian Coventry, et al. Sequence-specific targeting of intrinsically disordered protein regions. *bioRxiv*, pp. 2024–07, 2024.

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. *arXiv preprint arXiv:2302.02277*, 2023.

Vinicius Zambaldi, David La, Alexander E Chu, Harshnira Patani, Amy E Danson, Tristan OC Kwan, Thomas Frerix, Rosalia G Schneider, David Saxton, Ashok Thillaisundaram, et al. De novo design of high-affinity protein binders with alphaproteo. *arXiv preprint arXiv:2409.08022*, 2024.

APPENDIX

A SUPPLEMENTARY DATASET INFORMATION

A.1 PPIBENCH CURATION DETAILS

We curate a large-scale general protein-protein complex dataset by identifying chain-pair interfaces, following the data processing pipeline introduced in AlphaFold3 (Abramson et al., 2024). The curation process begins by applying a series of quality filters to identify valid PDB entries:

- Retain structures with a reported resolution of 9Å or better.
- Remove polymer chains containing unknown residues.
- Exclude protein chains where consecutive C_α atoms are separated by more than 10Å.
- Filter out polymer chains with fewer than four resolved residues.
- Eliminate clashing chains, defined as chains with more than 30% of their atoms located within 1.7Å of an atom in another chain. When two chains clashed:
  - The chain with the higher percentage of clashing atoms is removed.
  - If the clashing percentage is equal, the chain with fewer total atoms is removed.
  - If both chains have the same number of atoms, the chain with the larger chain ID is removed.

This filtering process result in 68,891 valid PDB entries. From these entries, chain-pair interfaces are identified, defined as pairs of chains with a minimum heavy atom separation of less than 5Å. Furthermore, to ensure efficient processing, all complexes are filtered to exclude sequences longer than 1,024 residues, finally resulting in a total of 301,182 chain-pair interfaces.

A.2 PPIBENCH DATA STATISTICS

	Complex
Training	602,129
Validation	189
Test	46
Total	602,364

Table 4: Data statistics for PPIBench.

We provide the data statistics for the curated PPIBench in Table 4.

A.3 DATA STATISTICS FOR TARGET-PROTEIN MINI-BINDER COMPLEX DATASET

Models	Seen Class					Zero-Shot				
	FGFR2	InsulinR	PDGFR	TGFb	VirB8	H3	IL7Ra	EGFR	TrkA	Tie2
Training	465	192	210	77	57	–	–	–	–	–
Validation	57	23	26	9	7	–	–	–	–	–
Test	57	23	26	9	7	38	7	4	4	2
Total	579	238	262	95	71	38	7	4	4	2

Table 5: Detailed data statistics for curated target-protein mini-binder complex dataset.

We collect experimentally confirmed positive target-protein mini-binder complexes against ten targets with diverse structural properties from Bennett et al. (2023). The detailed data statistics for the ten target proteins are reported in Table 5.

A.4 Data Statistics for Antigen-Antibody Complex Dataset

	CDR-H1 Clustering	CDR-H2 Clustering	CDR-H3 Clustering
Complexes	4,261	4,261	4261
Clusters	1,034	1,418	2,254
Training	3,590	3,576	3,427
Validation	403	399	451
Test	268	286	383

Table 6: Detailed data statistics for antigen-antibody complex dataset.

We detail the data statistics of antigen-antibody complex dataset for each CDR clustering in Table 6.

B Supplementary Implementation Details

Diffusion Schedules. Continuous spatial noise scheduling is parameterized across both Euclidean and rotational manifolds. For translation vectors in $\mathbb{R}^3$, the variance schedule scales from $\sigma_{\min} = 0.1$ to $\sigma_{\max} = 20.0$. We use a coordinate scaling factor of 0.1. For rotation matrices in $SO(3)$, the schedule scales from $\sigma_{\min} = 0.1$ to $\sigma_{\max} = 1.5$.

Architecture & Hyperparameters. The Invariant Point Attention (IPA) modules are configured with 8 attention heads, utilizing 8 query/key points and 12 value points per head. Loss hyper-parameters are set to $\alpha = 1.0$, $\gamma = 1.0$, and $\omega = 0.25$, and the threshold τ for the pairwise distance constraint is set to 0.25.

Training Setup. We train PPFrame on a single NVIDIA B200 GPU for up to 1.2 million steps. Optimization is performed with a learning rate of $1\mathrm{e}{-4}$ and a batch capacity of $1{,}536$ tokens per step. Validation is executed per $5,000$ gradient steps.

C Supplementary Experimental Results

C.1 Structure Diversity on the General Protein-Protein Complex Design Task

	Structure Diversity (↑)
RFD+ProteinMPNN	0.3097
APM	0.3063
Proteína-Complexa	0.1597
PPFrame	**0.3135**

Table 7: Structure diversity on the general protein-protein complex design task, which is quantified as $1 - \mathrm{TM\text{-}score}$ averaged across generated candidate structures.

Structural design diversity is quantified as $1 - \mathrm{TM\text{-}score}$ averaged across generated candidate structures. We interpret these diversity scores across three regimes: values above 0.5 represent high overall structural diversity; scores between 0.3 and 0.5 correspond to structures within the same superfamily that feature distinct loop configurations or subdomain topologies; and scores below 0.3 denote closely related structures sharing nearly identical secondary structure cores. As reported in Table 7, PPFrame achieves the highest structural diversity score, demonstrating a capacity to sample diverse loop structures and topological variations rather than collapsing onto rigid backbone templates.

C.2 Detailed Target-Protein Mini-Binder Complex Design Results

The detailed target-protein mini-binder complex design results for different targets are provided in Table 8 - 17.

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.637	78.92	36.84%	31.97%	68.25%	-42.04	63.15%	63.15%
APM	0.538	76.39	29.82%	29.05%	48.39%	-37.02	71.92%	54.38%
Complexa	0.823	80.67	54.38%	**49.25%**	76.91%	-45.81	91.22%	73.68%
PPFrame	**0.931**	**82.58**	**61.40%**	41.75%	**77.74%**	**-50.87**	**100%**	**98.24%**

Table 8: Model performance on target-protein mini-binder complex design for target FGFR2. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026). PPFrame achieves the best performance across sequence evaluation, structural evaluation and sequence-structure consistency metrics. Additionally, PPFrame also obtains the highest novelty score. Altogether, the results validate PPFrame is capable of designing high-affinity and novel protein complexes.

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.789	80.43	56.52%	30.89%	69.18%	-35.35	65.21%	65.21%
APM	0.562	79.01	43.47%	26.91%	52.91%	-31.04	78.26%	82.61%
Complexa	0.832	86.90	78.26%	**41.93%**	76.93%	-41.32	100%	86.95%
PPFrame	**0.895**	**93.04**	**86.95%**	27.54%	**79.51%**	**-46.92**	**100%**	**100%**

Table 9: Model performance on target-protein mini-binder complex design for target InsulinR. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026). PPFrame achieves the highest sequence design success rate, structure binding energy and sequence-structure co-designability, validating its ability in simultaneously designing high-quality and consistent protein-protein complex sequences and structures.

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.598	76.56	3.84%	28.93%	72.83%	-59.37	53.84%	53.84%
APM	0.391	73.89	0.0	30.27%	52.19%	-61.54	65.38%	61.53%
Complexa	0.652	78.09	**19.23%**	**41.25%**	79.42%	-66.56	88.46%	**88.46%**
PPFrame	**0.678**	**81.73**	7.69%	25.78%	**82.43%**	**-68.82**	**100%**	30.76%

Table 10: Model performance on target-protein mini-binder complex design for target PDGFR. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026). PPFrame achieves better structure design performance, but lags behind Proteína-Complexa on both sequence design quality and sequence-structure design consistency.

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.589	76.90	11.11%	35.17%	68.23%	-31.09	55.56%	55.56%
APM	0.461	72.10	11.11%	32.81%	48.65%	-23.16	77.78%	11.11%
Complexa	0.683	**80.15**	22.22%	46.17%	76.12%	-46.54	88.89%	**33.33%**
PPFrame	**0.733**	78.17	**22.22%**	**48.57%**	**76.46%**	**-49.29**	**100%**	22.22%

Table 11: Model performance on target-protein mini-binder complex design for target TGFb. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026). PPFrame achieves better performance on both sequence and structure design quality. However, the sequence-structure co-designability of PPFrame is a bit lower than that of Proteína-Complexa.

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.570	82.89	14.28%	32.95%	72.91%	-32.90	57.14%	57.14%
APM	0.368	85.09	0.0	30.19%	52.87%	-19.59	71.74%	42.85%
Complexa	0.602	86.10	14.28%	**46.52%**	80.15%	-38.56	85.71%	85.71%
PPFrame	**0.657**	**86.67**	**14.28%**	36.17%	**81.16%**	**-48.31**	**100%**	**100%**

Table 12: Model performance on target-protein mini-binder complex design for target VirB8. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026). PPFrame obtains better sequence and structure design quality as well as sequence-structure consistency.

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.571	83.09	**25.00%**	29.15%	71.26%	-46.87	25.00%	25.00%
APM	0.512	83.19	0.0	27.93%	51.34%	-32.76	50.00%	50.00%
Complexa	0.627	86.76	0.0	**40.16%**	76.31%	-40.90	100%	50.00%
PPFrame	**0.680**	**90.19**	0.0	22.62%	**82.56%**	**-48.35**	**100%**	**100%**

Table 13: Model performance on target-protein mini-binder complex design for target EGFR. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026). PPFrame achieves better structure binding energy and sequence-structure co-designability. However, the sequence design success rate of PPFrame is zero, which might be due to the lack of deep pockets of EGFR targets and the zero-shot setting of this class.

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.782	78.56	52.63%	28.16%	71.04%	-35.93	26.31%	26.31%
APM	0.562	76.29	47.38%	23.91%	53.16%	-36.45	49.87%	42.10%
Complexa	0.891	79.05	60.52%	**38.16%**	80.47%	-40.65	78.93%	**47.36%**
PPFrame	**0.934**	**82.32**	**97.36%**	21.97%	**83.16%**	**-47.08**	**100%**	31.57%

Table 14: Model performance on target-protein mini-binder complex design for target H3. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026). PPFrame obtains higher sequence design success rate and lower structure binding energy, confirming that PPFrame has the capability to design high-affinity binders even in zero-shot setting.

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.543	85.67	14.28%	29.18%	70.19%	-40.08	42.85%	42.85%
APM	0.429	78.09	0.0	27.56%	53.96%	-23.93	85.71%	57.14%
Complexa	0.634	89.90	28.57%	**36.17%**	80.16%	-42.26	100%	71.42%
PPFrame	**0.734**	**92.21**	**42.85%**	20.35%	**82.30%**	**-50.42**	**100%**	**100%**

Table 15: Model performance on target-protein mini-binder complex design for target IL7Ra. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026). PPFrame achieves higher sequence design success rate, lower structure binding energy and higher sequence-structure co-designability. Note that the designability and co-designability of PPFrame on this target are both 100%, validating the strong binder design capability of PPFrame in the zero-shot setting.

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.562	86.90	**50.00%**	29.31%	72.91%	-38.95	75.00%	75.00%
APM	0.421	82.19	0.0	30.19%	51.43%	-27.85	100%	50.00%
Complexa	0.590	90.10	0.0	**39.42%**	81.58%	-33.83	100%	50.00%
PPFrame	**0.635**	**92.98**	0.0	10.52%	**88.20%**	**-42.39**	**100%**	**100%**

Table 16: Model performance on target-protein mini-binder complex design for target Tie2. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026). PPFrame achieves higher structure design quality and sequence-structure consistency.

Models	Sequence					Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	Diversity (↑)	Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.791	72.08	25.00%	28.17%	70.16%	-33.28	50.00%	50.00%
APM	0.563	73.90	25.00%	31.42%	49.87%	-31.92	75.00%	25.00%
Complexa	0.829	80.10	50.00%	**47.92%**	78.93%	-41.50	100%	50.00%
PPFrame	**0.840**	**80.31**	**75.00%**	35.49%	**78.92%**	**-46.04**	**100%**	**75.00%**

Table 17: Model performance on target-protein mini-binder complex design for target TrkA. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026). Similarly, PPFrame achieves better sequence design success rate, structure binding energy and sequence-structure consistency, confirming its ability in designing good binders in the zero-shot setting.

C.3 Detailed Antigen-Antibody Complex Design Results for Each CDR Clustering

We provide the detailed results of antigen-antibody complex design for each CDR clustering in Table 18 - 20.

Models	Sequence						Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	H1-Novelty (↑)	H2-Novelty (↑)	H3-Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.538	80.67	15.67%	49.78%	51.26%	59.31%	-30.03	58.20%	58.20%
APM	0.239	79.23	7.46%	42.18%	53.37%	63.17%	-29.38	57.08%	52.61%
Complexa	0.523	81.09	14.92%	55.36%	**68.26%**	72.93%	-35.35	88.43%	90.67%
PPFrame	**0.548**	**82.10**	**17.16%**	**56.64%**	66.46%	**74.68%**	**-42.49**	**100%**	**100%**

Table 18: Model performance on antigen-antibody complex design task for CDR-H1 clustering. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026).

Models	Sequence						Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	H1-Novelty (↑)	H2-Novelty (↑)	H3-Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.497	79.29	17.48%	43.87%	58.69%	62.89%	-32.98	58.04%	58.04%
APM	0.271	76.32	5.59%	45.28%	52.97%	63.16%	-28.78	52.98%	53.14%
Complexa	**0.519**	**81.90**	17.83%	**52.79%**	65.23%	**70.95%**	-30.13	91.25%	93.35%
PPFrame	0.506	80.62	**19.58%**	49.38%	**67.21%**	69.61%	**-44.37**	**100%**	**100%**

Table 19: Model performance on antigen-antibody complex design task for CDR-H2 clustering. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026).

Models	Sequence						Structure		Co-designability (↑)
	ipTM (↑)	pLDDT (↑)	SR (↑)	H1-Novelty (↑)	H2-Novelty (↑)	H3-Novelty (↑)	ΔG (↓)	Designability (↑)	
RFD+IF	0.487	80.17	17.23%	48.26%	58.73%	63.75%	-36.25	61.62%	61.62%
APM	0.259	78.23	6.78%	49.67%	56.19%	66.97%	-21.89	50.65%	52.48%
Complexa	0.502	80.26	18.01%	**53.26%**	65.29%	**71.86%**	-32.90	91.12%	92.95%
PPFrame	**0.525**	**81.50**	**19.58%**	50.88%	**69.85%**	71.51%	**-46.53**	**95.03%**	**98.17%**

Table 20: Model performance on antigen-antibody complex design task for CDR-H3 clustering. (↑): the higher the better. (↓): the lower the better. IF denotes the inverse folding model ProteinMPNN (Dauparas et al., 2022) and Complexa denotes Proteína-Complexa (Didi et al., 2026).

C.4 Detailed Binder Design Performance on the Pretrained Model

Performance metrics for the pretrained PPFrame on target-protein mini-binder complex design are presented in Table 21. Notably, even without domain-specific adaptation, PPFrame exhibits robust zero-shot capabilities across diverse targets, most prominently in sequence-structure consistency, where it achieves 100% co-designability across all evaluated targets. Subsequent supervised fine-tuning on a curated mini-binder dataset further elevates overall design quality, markedly boosting the mean sequence design success rate from 8.72% to 40.77%. These results confirm that while pretraining imparts strong zero-shot structural alignment, targeted fine-tuning substantially enhances sequence-level binding optimization.

Models	FGFR2 (Seen)			InsulinR (Seen)			PDGFR (Seen)		
	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)
PPFrame	**61.40%**	-50.87	98.24%	**86.95%**	**-46.92**	**100%**	**7.69%**	-68.82	30.76%
w/o finetuning	5.52%	**-55.88**	**100%**	8.69%	-44.08	100%	0.0	**-72.28**	**100%**
Models	TGFb (Seen)			VirB8 (Seen)			EGFR (Zero-shot)		
	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)
PPFrame	**22.22%**	**-49.29**	22.22%	**14.28%**	**-48.31**	**100%**	0.0	**-48.35**	**100.0%**
w/o finetuning	0.0	-42.49	**100%**	14.28%	-41.87	100%	0.0	-45.92	100%
Models	H3 (Zero-shot)			IL7Ra (Zero-shot)			Tie2 (Zero-shot)		
	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)
PPFrame	**97.36%**	-47.08	31.57%	**42.85%**	**-50.42**	**100%**	0.0	**-42.39**	**100.00%**
w/o finetuning	44.73%	**-52.20**	**100%**	14.27%	-47.12	100%	0.0	-39.38	100%
Models	TrkA (Zero-shot)			Average			Novelty (↑)	Diversity (↑)	
	SR (↑)	ΔG (↓)	Consistency (↑)	SR (↑)	ΔG (↓)	Consistency (↑)			
PPFrame	**75.00%**	-46.04	75.00%	**40.77%**	**-49.85**	75.78%	81.24%	**29.08%**	
w/o finetuning	0.0	**-48.55**	**100%**	8.72%	-48.98	**100%**	**83.32%**	28.67%	

Table 21: Pretrained model performance on target-protein mini-binder complex design task. (↑): the higher the better. (↓): the lower the better.