

Unlocking diverse molecular binding modes using latent feature sampling

Houtim Lai

mosquitolkfo@gmail.com

Tencent <https://orcid.org/0009-0005-2507-465X>

Peng Zhou

Hunan University <https://orcid.org/0000-0001-9342-6024>

Huaxian Jia

Tencent AI for Life Sciences Lab

Kun Xie

Tencent

Geyan Ye

Tencent

Wei Liu

Tencent AI lab, Tencent

Houtim Lai

Tencent AI for Life Sciences Lab

Xiangxiang Zeng

Hunan University <https://orcid.org/0000-0003-1081-7658>

Article

Keywords: Protein structure prediction, Binding site identification, Protein-protein docking, Antibody-antigen docking, Protein-DNA interactions, Latent feature analysis

Posted Date: April 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9511630/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Unlocking diverse molecular binding modes using latent feature sampling

Houtim Lai^{1*†}, Peng Zhou^{2,1†}, Huaxian Jia^{1*}, Kun Xie¹,
Geyan Ye¹, Wei Liu¹, Alexander Ng¹, Xiangxiang Zeng^{2*}

¹AI for Life Sciences Lab, Tencent, Shenzhen, China.

²College of Computer Science and Electronic Engineering, Hunan
University, Hunan, China.

*Corresponding author(s). E-mail(s): mosquitolkfo@gmail.com;
huaxianjia@tencent.com; xzeng@hnu.edu.cn;

[†]These authors contributed equally to this work.

Abstract

Deep learning has revolutionized the prediction of biomolecular structures, yet accurately identifying diverse intermolecular binding sites remains a persistent challenge. Current state-of-the-art models often stagnate in dominant local minima determined by co-evolutionary biases, failing to sample biologically critical alternative conformations. Here we present DynaHelix, a framework that actively steers the generative process to unlock diverse binding modes encoded within the latent space of structure prediction models. Unlike methods relying on stochastic perturbations of multiple sequence alignments, DynaHelix analyzes intermediate pairwise feature to identify potential interaction hotspots. These are converted into targeted geometric constraints that drive the diffusion model to explore specific, under-represented interfaces. We demonstrate that DynaHelix significantly outperforms leading open-source models including Protenix, Chai-1, and Boltz-1 across comprehensive benchmarks FoldBench and challenging targets from CASP15 and CASP16. Notably, our approach achieves superior structural accuracy for complex antigen–antibody and protein–DNA interactions while reducing the computational sampling budget by an order of magnitude. By enabling deterministic control over the sampling landscape, DynaHelix offers a generalizable pathway for resolving high-ambiguity molecular interactions essential for mechanistic biology and drug discovery.

Keywords: Protein structure prediction, Binding site identification, Protein-protein docking, Antibody–antigen docking, Protein–DNA interactions, Latent feature analysis

1 Introduction

Interactions between molecules are fundamental to understanding the mechanisms of life and are central to drug discovery and biotechnology. While recent deep learning-based structure prediction tools show superior performance in predicting structures from sequences [1–4], the principles governing the identification of binding interfaces and their sampling space remain elusive. Previous studies suggest that Multiple Sequence Alignments (MSAs) provide a strong initial bias, constraining sampling to a narrow region around this initial guess [5]. Consequently, a key limitation of these models is that their sampling strategy often stagnates in local minima corresponding to different interfaces, rather than exploring potential alternative binding modes to find the global minimum, as conceptually illustrated in Fig. 1a. This inefficiency poses a prohibitive computational burden. For instance, in the Critical Assessment of Techniques for Protein Structure Prediction (CASP) 15 [6], AFsample [7, 8] required 6,000 sampling iterations for target H1144 to retrieve only five high-quality (high DockQ [9]) samples, highlighting the inaccessibility to these binding modes to current methods. Similarly, in CASP16 Phase 2 [10, 11], where the committee provided over 8,040 predicted structures per participating group [12], a lack of structural diversity represents a substantial waste of computational resources. Therefore, developing methods to retrieve an equivalent set of representative structures with significantly fewer sampling iterations is essential for optimizing computational efficiency.

Recent open-source state-of-the-art (SOTA) structure prediction models, including Chai-1 [13], Boltz-1 [14], and Protenix [15], adopt core architectures analogous to AlphaFold3. These frameworks typically take molecular sequences and associated MSAs as input, processing them through a pairformer module to yield a pairwise feature z , which is subsequently utilized by a diffusion module to generate predicted structures. To induce structural diversity within this deterministic framework, strategies such as network dropout and MSA column-wise masking have been employed to increase uncertainty, perturb the initial guess [7, 16]. While such perturbations When combined with extensive sampling in AlphaFold2 [17], these perturbations can significantly improve predictive accuracy. Additionally, research on MSA engineering indicates that techniques such as sub-sampling, mutations, and clustering can enhance the diversity of the sampling space [18–23]. Furthermore, gradient descent methods have been used to denoise subsampled MSA, helping the model converge to maximize confidence scores [24]. However, these methods have significant limitations. Adding dropout can result in the loss of core features, while offering only limited diversity enhancement, and massive sampling is computationally inefficient. Moreover, since MSA engineering relies heavily on the depth and coverage of available MSAs, there remains no straightforward or intuitive strategy to effectively enhance diversity without these drawbacks.

To overcome these MSA-dependent bottlenecks, one might look to alternative sampling paradigms that bypass co-evolutionary signals entirely. Traditional protein–protein docking tools, such as ClusPro, HADDOCK and RosettaDock [25–27], utilize Fast Fourier Transform (FFT) [28] to systematically explore the entire binding landscape. However, these methods are computationally demanding and largely restricted to rigid-body docking, failing to account for the essential flexibility of protein

interfaces. A more promising avenue emerges from recent architectures like Chai-1 and Protenix that introduce a direct control mechanism via explicit geometric constraints. In principle, such constraints can steer the generation process toward user-specified pockets or specific molecular interactions, rendering structure prediction controllable. However, a significant practical challenge persists because accurate constraints are rarely known a priori and there is currently no standard methodology to reliably retrieve or predict them.

We present an intuitive method to systematically govern the initial sampling guess by capitalizing on the latent structural hypotheses encoded within the model itself. While MSAs provide valuable evolutionary context, our approach advances beyond reliance on their stochastic variation to drive diversity. By mining the latent pairwise feature to identify high-probability interaction hotspots, we automatically generate potential interface (PI) constraints. These constraints exert a targeted driving force conceptually illustrated in Fig. 1a that actively steers the sampling trajectory away from initial evolutionary biases and toward the global minimum near the ground truth structure. This explicit steering mechanism enables the discovery of novel binding modes and enhances generality to unseen targets, achieving superior structural diversity with significantly reduced computational cost.

2 Model

Here, we present DynaHelix, a novel deep learning framework for multi-molecule structure prediction capable of automatically sampling key binding modes encoded within the pairwise feature space z . Building upon an architecture analogous to AlphaFold3, DynaHelix integrates a specialized Potential Interface Finder (PIF) algorithm. This module analyzes the z feature to identify critical cross-chain contact patches between the receptor and binder, which are subsequently converted into interface constraints. These constraints are fed back into the input embedder, effectively steering the model to focus its sampling on specific receptor pockets and generating distinct, interface-specific conformations, as outlined in Fig. 1b.

A central challenge in this field is the unsupervised identification of Potential Interfaces (PIs). We observed that the predicted z feature implicitly encodes a superposition of multiple binding modes rather than a single specific interface (section 3.1). To address this, we developed the PIF algorithm, which first ranks interaction hotspots based on cross-chain attention weights. It then identifies spatially coherent clusters of receptor tokens that exhibit similar binding patterns. Similar interfaces are clustered and representative centers are selected, yielding a top- k set of distinct interfaces that guide the model to generate corresponding interface-specific structures (Fig. 1c, section 3.2).

Since standard AlphaFold3-like architectures lack a native mechanism for user-directed interface generation, we introduce a constraint model that steers sampling toward the target sampling region based on user-defined constraints. To enable this, a constraint embedder is trained to learn representations from randomly sampled contact pairs extracted from ground-truth structures, allowing the model to

interpret and adhere to spatial constraints during inference. (Fig. 1d, supplementary section B.1).

In the following sections, we first demonstrate that the pair feature z governs the sampling space and reveals the multiple binding modes hidden within it. We then optimize key PIF parameters and analyze their impact on performance metrics. Subsequently, we present detailed case studies to visualize the algorithm’s efficacy. Finally, we provide a quantitative benchmarking of DynaHelix, demonstrating its superior performance across diverse hetero-complexes, including antigen-antibody, protein-protein, and protein-dsDNA (double-stranded DNA) interactions [29–31].

3 Results

3.1 The pair feature z governs the sampling space

The pair feature z is derived from the PairFormer using input sequences and MSAs, and governs the sampling space by encoding token-level pairwise hidden states. Although the final structures are realized by the downstream structure module, the integrity of z is critical for valid generation. Previous work by Hayes et al. demonstrated that replacing z with zero or randomized tensors induces structural collapse, rendering the model unable to generate valid structures [32]. While methods such as Entropy Guided Fold utilize an auxiliary loss function to modulate z to reduce structural bias in AlphaFold2 [33], and Boltz-sample employs a global scalar to enhance the recovery of monomeric multi-states [34], we hypothesize that directly modifying z can deterministically steer the model toward designed binding interfaces. We further propose that using constraint embeddings significantly facilitates this goal. To elucidate the underlying mechanism, we investigated its behavior based on z across three distinct molecular interaction types: antibody-antigen (PDB: 7R58), protein-protein (PDB: 7UPM), and protein-dsDNA (PDB: 7EDS) [35–37].

The first row of Fig. 2a (i-iii) displays the raw pair embeddings z , where regions of high intensity represent the most probable binding modes. Notably, a single seed with a fixed MSA can reveal multiple PIs. In the case of 7R58, sampling was distributed between two dominant interfaces, whereas for 7UPM and 7EDS, the sampled structures aligned primarily with the single dominant weighted regions (PI-1 and PI-2, respectively). To determine if the magnitude of these weights directly dictates sampling, we manually perturbed the embeddings by adding a scalar value of 50 to the PI-1 and PI-2 regions (Row 2). As expected, increasing the weight of PI resulted in the generated conformations having contact surfaces that align with PI. However, the resulting structures frequently exhibited steric clashes or poor interface quality. This suggests that while magnitude directs the location of sampling, the downstream structure module struggles to resolve the geometry of these artificially edited features, likely because such distributions were not encountered during training. Finally, we examined the use of constraints to guide sampling, a mechanism employed by models such as Chai-1 and Protenix (Row 3). Applying constraints yielded embeddings similar to the manually edited versions but with coherent adjustments to the surrounding feature space. Unlike manual editing, this constraint-based approach produced physically

valid structures with refined binding interfaces, evidenced by improved DockQ scores, that strictly adhered to the target PIs across all three cases.

To quantitatively compare the different pair embeddings, we calculated the average normalized weight across all binder tokens for each respective receptor token. These averaged token weights are presented as a line plot in Fig. 2b, where receptor tokens corresponding to the designed PI region are highlighted in bold red along the x-axis. The edited pair embedding (green triangles) exhibits significant boosting of weights within the designated PI region while maintaining consistency with the raw pair embedding (blue circles) in the non-PI regions. Conversely, while the constraint embedding (red stars) similarly enhances weights in the target PI region, it also demonstrates coherent adjustments to the weights in the surrounding non-PI regions relative to the baseline. This analysis provides further evidence that the constraint mechanism induces a more coherent and naturally distributed feature space, enabling the downstream structure module to adaptively predict well-refined structures of the complex.

3.2 Automated discovery of potential binding interfaces

In practical scenarios where the ground truth binding site is unknown, accurate interface identification is a prerequisite for successful complex structure prediction. Building on the insights from section 3.1, we developed the PIF to autonomously extract a finite set of high-probability binding interfaces directly from the pair feature z . As detailed in methods section 5.3, the PIF algorithm systematically identifies potential interfaces, iteratively designating chains as the binder and ranking PIs by their predicted weights. Each predicted interface defines a specific binder and a constituent set of receptor tokens, thereby delineating the key interacting residues. The efficacy of this automated search strategy relies on two governing hyperparameters: $\tau \in [0, 1]$, which establishes a binding interface size threshold, and the sampling depth k , which determines the number of top-ranked interfaces retained. Increasing k expands the search space, theoretically improving the likelihood of capturing the native interface, though at a higher computational cost.

To evaluate this approach, we employed a low-homology validation dataset comprising targets released after the training cutoff (supplementary section A.1). Fig. 3 characterizes the interface discovery performance across three distinct interaction modalities. The “Hit Rate” (Fig. 3, column 1) quantifies the algorithm’s ability to locate the correct binding pocket, defined as the retrieval of interfaces where at least one predicted receptor token lies within 5 Å of the ground truth binder. We observed a distinct gradient in which top-ranked interfaces (low k) exhibit significantly higher accuracy than lower-ranked ones, validating the efficacy of the scoring function in prioritizing native-like signals. Notably, the hit rate remains robust across varying τ values, suggesting that the ranking mechanism is stable and less sensitive to the specific inclusion threshold than to the rank order. Among the modalities, Protein-dsDNA complexes (Row c) presented the strongest structural signals, followed by Protein-Protein complexes (Row b), while Antibody-Antigen complexes (Row a) proved the most challenging due to the hypervariability of the paratope regions. We further assessed the signal purity of the extracted constraints via “Interface Precision” (Fig. 3,

column 2), which measures the proportion of true contacting tokens within a predicted interface. As the strictness of the threshold τ increases, the precision of the identified contacts improves markedly, demonstrating that higher confidence cutoffs effectively filter out background noise to provide cleaner guidance signals for the structural module. Consistent with the Hit Rate analysis, the highest-ranked predictions yielded the highest fidelity, reinforcing the reliability of the PIF’s prioritization. The impact of these search parameters on structural accuracy is shown in the DockQ scores for the “Oracle” (best sampled structure) and “Top-1” scenarios (Fig. 3, columns 3 and 4). While expanding the search depth (k) initially improves the Oracle performance by broadening the conformational search space, the benefits plateau around $k = 5$. Beyond this point, further sampling yields diminishing returns, indicating that the top 5 candidates sufficiently capture the relevant functional subspace. Consequently, we fixed $k = 5$ for subsequent analyses to maintain an optimal balance between prediction accuracy and computational efficiency.

Finally, the analysis reveals that the optimal search parameters are dictated by the interaction modality, mirroring the distinct physicochemical architectures of the binding interfaces. Antibody-antigen complexes required the most stringent filtering ($\tau = 0.9$), likely because these interactions are mediated by hypervariable complementarity-determining regions (CDRs), where high-affinity binding relies on a concentrated set of key residues (hotspots) within a structurally constrained paratope [38, 39]. Such architectures necessitate stricter thresholds to distinguish true signals from background noise. In contrast, intermediate thresholds proved sufficient for the other modalities ($\tau = 0.6$ for protein-protein and $\tau = 0.5$ for protein-dsDNA), consistent with their broader interaction footprints. General protein-protein interactions typically involve shape-complementary surfaces stabilized extensively by the burial of hydrophobic patches [40, 41], while protein-dsDNA complexes rely on extensive contact surfaces to read base sequences within the major and minor grooves [42, 43]. These widespread footprints yield robust geometric signals that are readily detected even at lower confidence levels.

3.3 DynaHelix overcomes sampling traps in complex energy landscapes

To empirically validate the efficacy of our DynaHelix framework, we revisited the three challenging case studies introduced in section 3.1. We performed a comparative analysis between DynaHelix and the Protenix base model (training details in supplementary section B). Both models were subjected to a standardized inference protocol, generating 125 candidate structures per target (details in supplementary section A.4) to ensure a rigorous assessment of sampling capability.

Fig. 2c details the modeling landscapes for these distinct interaction modalities. The scatter plots (top panels) delineate the relationship between the model’s self-assigned confidence score and structural accuracy (DockQ). The base model (black points) consistently fails to locate the native binding mode, producing only non-native decoys with DockQ scores near zero. In contrast, DynaHelix (red points) successfully samples high-accuracy conformations. Crucially, the model’s confidence head generally

assigns higher ranking scores to these near-native structures, facilitating their distinction from incorrect decoys. This indicates that the PIF-derived constraints guide the sampling process towards native-like basins, thereby enhancing the model’s ability to prioritize correct binding modes within the sampled ensemble.

To further investigate the sampling landscape behavior, we projected the conformational ensemble into a 2D latent space via Principal Component Analysis (PCA) and overlaid an energy landscape calculated using Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) [44, 45]. By integrating gas-phase molecular mechanics energies with an implicit continuum solvation model, MM/GBSA provides a robust approximation of the binding free energy (ΔG_{bind}), detailed in supplementary section B.4. This metric quantifies the thermodynamic stability of the binder-receptor interaction, serving as a physics-based validator independent of the deep learning potentials. The resulting landscapes (Fig. 2c, bottom panels) reveal that the base model exhibits restricted sampling capabilities, frequently becoming trapped in local energy minima. For instance, in the antibody-antigen case (7R58), the base model explores PI-1 and PI-2 but completely fails to access the native-like PI-3 region. Similarly, in 7UPM and 7EDS, the base model is heavily biased toward a specific non-native single region (PI-2), resulting in a narrow distribution of incorrect structures.

Conversely, DynaHelix demonstrates a comprehensive exploration of the free energy landscape. By integrating potential interface constraints, the model successfully traverses energy barriers to populate all three prioritized binding modes. As visualized by the centroid structures (Fig. 2c, right), when the searched interface overlaps with the ground truth, DynaHelix resolves the complex with high structural fidelity, effectively escaping the local minima that confound the baseline approach.

3.4 Diverse interface sampling drives performance gains across comprehensive benchmarks

To rigorously evaluate DynaHelix, we benchmarked its performance against three state-of-the-art open-source structure models: Protenix, Chai-1, and Boltz-1 across three datasets: an internal validation set, FoldBench [46], and CASP 15/16 [6, 10]. To ensure a fair comparison, we enforced computational parity by standardizing the sampling budget to 125 structures per target across all methods. Specifically, as DynaHelix samples across five distinct predicted interfaces, we adjusted the baseline generation protocols to match this total volume (detailed in supplementary section A.4). We further introduced IfaceRanking (method section 5.4), a hierarchical selection metric that identifies the optimal structure for each interface and ranks them to determine the final top- k predictions. The following analysis reports performance based on the Top-5 scenarios using this protocol.

Internal Validation: Isolating algorithmic contributions. To strictly evaluate the contribution of DynaHelix, we utilized a held-out validation set of low-homology targets released after the training cutoff. We benchmarked DynaHelix against two internal baselines: a Base model (unconstrained sampling) and a Site-Oracle (sampling with constraints derived from the four closest ground-truth receptor tokens). The Site-Oracle serves as a strong upper bound for local contact guidance.

As illustrated in Fig. 4a, the antibody-antigen category presented the greatest challenge. While the Base model struggled (0.160 DockQ), the Site-Oracle achieved 0.515, highlighting that accurate structures are generated if the interface is localized. DynaHelix significantly bridged this gap, improving upon the Base model by +0.110 (0.160 to 0.270). Conversely, for protein-protein and protein-dsDNA complexes, the Base model performance was already robust, resulting in narrower margins of improvement (+0.11 and +0.03, respectively). Notably, however, DynaHelix outperformed the Site-Oracle upper bound in both categories: +0.034 in protein-protein (0.557 vs. 0.523) and +0.037 in protein-dsDNA (0.696 vs. 0.659). We hypothesize that while Site-Oracle relies on rigid constraints from a localized patch (the closest tokens), DynaHelix identifies spatially distinct contacts across the binding surface. These dynamic, multi-modal constraints likely facilitate superior global structural accommodation compared to rigid local proximity alone.

FoldBench: Robustness on unseen targets. To assess generalization, we evaluated performance on FoldBench [46], a comprehensive dataset released in late 2025 comprising recent PDB depositions. DynaHelix consistently surpassed all three SOTA baselines across Oracle, Top-5, and Top-1 metrics. The performance advantage was most pronounced in the antibody-antigen category (Fig. 4b). Specifically, DynaHelix-IfaceRanking achieved a Top-5 score of 0.441, outperforming the second-best method, Protenix (0.352), by a substantial margin of +0.089. Given the scale and currency of this benchmark, these results underscore the method’s capacity to generalize effectively to unseen targets.

CASP 15/16: Efficiency in high-difficulty regimes. We further tested DynaHelix on critical targets from CASP15 (2022) and CASP16 (2024), representing high-difficulty benchmarks released after the training cutoff. The evaluation set included 11 hetero-complexes from CASP15 and 6 antibody-antigen targets from CASP16 (n=17, detailed in supplementary section A). Unlike typical CASP protocols that utilize massive sampling (often > 1,000 seeds) [12], we restricted the budget to 100 random seeds to rigorously evaluate sampling efficiency (Fig. 4c). Under this constrained budget, DynaHelix established a commanding lead. In the Oracle metric, our method reached a DockQ score of 0.663, surpassing the runner-up, Protenix (0.530), by +0.133. In the practical Top-5 scenario, this gap widened further: DynaHelix-IfaceRanking achieved 0.551, outperforming Protenix (0.365) by +0.186. This disparity indicates that while current SOTA models struggle to resolve these challenging interfaces, DynaHelix successfully navigates the energy landscape. Finally, we analyzed sampling efficiency (Fig. 4c, right). Remarkably, DynaHelix requires only 5 seeds to match the performance of Protenix at 100 seeds. Furthermore, the performance curve for DynaHelix maintains a steep upward trajectory at 100 seeds, suggesting that additional compute would yield further structural refinements.

To provide a comprehensive qualitative assessment, Fig. 5a displays the structural predictions for all targets in the CASP dataset. In these visualizations, the best structure predicted by DynaHelix is rendered in color and superimposed onto the ground truth structure, which is shown in grey. A detailed quantitative breakdown of these results is presented in Fig. 5b, where individual target scores are categorized

by accuracy. We utilize a color-coded classification scheme to denote model performance: red indicates failed predictions, while yellow, green, and blue correspond to acceptable, medium, and high-accuracy predictions, respectively. Overall, DynaHelix demonstrates robust performance across the dataset. The only notable failure case is target H1129, where the model yielded a DockQ score below 0.23. Apart from this specific outlier, the method consistently generates structures that closely align with the ground truth.

4 Discussion

We have introduced DynaHelix, a deep learning framework designed to identify critical binding interfaces and improve structural prediction accuracy through enhanced sampling diversity. We demonstrated that the system effectively locates potential binding interfaces via the pairwise feature z . These identified interfaces are converted into constraint pairs that steer the model toward alternative binding modes, thereby preventing the model from getting trapped in local minima. Extensive evaluation across multiple benchmarks and diverse interaction types reveals DynaHelix significantly outperforms baseline models. Furthermore, we posit that the mechanism elucidated in this study can be adapted to other modern structure prediction architectures with minimal modification effort.

Despite these successes, our current approach utilizes a relatively coarse constraint format, defined by contacts between all binder tokens and specific receptor regions. Higher resolution constraints ranging from token-to-token to atom-to-atom interactions could theoretically be decomposed from the pairwise feature z . While implementing these granular constraints requires modification to the current method, they hold particular promise for improving Protein-Protein and Protein-dsDNA predictions, cases where the model often identifies the general binding interface but lacks precision in interfacial details.

A critical insight emerging from this study is the divergence between sampling capability and ranking accuracy in modern structure prediction models. While DynaHelix demonstrates a robust capacity to access near-native states, achieving a high Oracle DockQ of 0.663 on the challenging CASP dataset, the ranking accuracy lags behind, with a Top-1 score of 0.444. This divergence suggests that while the diffusion architecture can be effectively steered to generate diverse, physically valid binding modes, the model’s confidence head struggles to accurately distinguish the native solution from plausible decoys. We attribute this discrepancy to distributional shifts in the confidence head, which appears overly sensitive to the constraint features. Because the current training protocol relies on a limited number of static binding conformations, the confidence head is not exposed to diverse binding interfaces during training. Future directions should leverage variable interface constraints to generate diverse decoys, thereby training the confidence head on a more complete sampling space to narrow the Oracle-Top-1 gap.

Additionally, the current PIF relies on a rule-based heuristic algorithm. An important avenue for future research is the development of an end-to-end learnable module capable of automatically decomposing the pairwise feature z into potential interfaces.

Such an approach, while requiring significant modifications to the loss function and training protocols, would likely enhance model robustness by replacing fixed heuristics with data-driven feature extraction.

Finally, unlike traditional MSA engineering, DynaHelix offers a direct method for modulating the pairwise feature z . These results demonstrate a viable path for exploring the uncharted sampling space, potentially uncovering novel binding interactions and contributing valuable insights to the field of structural biology.

5 Methods

5.1 Preliminary

The structure prediction model maps primary sequences to 3D atomic coordinates through a hierarchical deep learning pipeline dependent on stochastic inputs. Formally, given an input sequence set $\mathcal{S}$ and a random seed ϵ , the inference workflow is expressed as a composition of functional modules:

$$\mathbf{Z}_\epsilon = \Psi_{\text{pair}}(\Omega_{\text{emb}}(\mathcal{S}, \epsilon)) \quad (1)$$

$$\mathbf{x}_\epsilon = \Phi_{\text{diff}}(\mathbf{Z}_\epsilon) \quad (2)$$

where Ω_{emb} denotes the Input Embedder. The seed ϵ governs the stochastic subsampling of MSA within Ω_{emb} , resulting in specific latent embeddings. These embeddings are processed by the pairformer Ψ_{pair} to generate a unique pairwise representation $\mathbf{Z}_\epsilon \in \mathbb{R}^{N \times N \times d}$ corresponding to the seed. Finally, the Diffusion Structure Module Φ_{diff} takes these representations to generate the atomic coordinates $\mathbf{x}_\epsilon$.

To characterize the interaction interface for a specific prediction, we analyze the contact probabilities and chain roles derived from this seed-specific pass. The distograms predicted by the model are converted into contact probability matrices $\mathbf{P}_d \in [0, 1]^{N \times N}$ at thresholds $d \in \{8, 20\}$ Å.

5.2 Constraint feature guides structure prediction

We formally define a geometric constraint set $\mathcal{K}$ as a tuple consisting of a target binder chain c_b and a selected subset of receptor token indices $\mathcal{I}_r$:

$$\mathcal{K} = (c_b, \mathcal{I}_r), \quad \text{where } c_b \in \mathcal{C}, \mathcal{I}_r \subset \{1, \dots, N\} \setminus \text{indices}(c_b) \quad (3)$$

The inference workflow is updated to incorporate these constraints via a residual connection. The Constraint Embedder, denoted as Ω_{const} , maps $\mathcal{K}$ into the same latent dimension as the sequence embeddings. The initial latent representation $\mathbf{H}$ is thus computed as the sum of the sequence and constraint embeddings:

$$\mathbf{H} = \Omega_{\text{emb}}(\mathcal{S}, \epsilon) + \Omega_{\text{const}}(\mathcal{K}) \quad (4)$$

Subsequently, this constraint-aware representation is propagated through the network to generate the final structure:

$$\mathbf{x}_{\epsilon, \mathcal{K}} = \Phi_{\text{diff}}(\Psi_{\text{pair}}(\mathbf{H})) \quad (5)$$

Concurrently, the complex is partitioned into a Binder set $\mathcal{B}$ and a Receptor set $\mathcal{R}$. This assignment is determined by the biological modality of the input complex using two distinct strategies:

- **Automatic Selection:** For general protein-protein and protein-dsDNA complexes, we adopt an exhaustive one-vs-rest approach. Every chain c in the complex $\mathcal{C}$ is iteratively designated as the binder, while the remaining chains serve as the receptor:

$$\forall c \in \mathcal{C} : \quad \mathcal{B} = \{c\}, \quad \mathcal{R} = \mathcal{C} \setminus \{c\} \quad (6)$$

- **Rule-based Selection:** For antibody-antigen complexes, we employ a deterministic assignment based on domain knowledge. Given that the heavy chain typically dominates the paratope-epitope interaction, we strictly designate the antibody heavy chain as the binder and the antigen chains as the receptor:

$$\mathcal{B} = \{c_{\text{heavy}}\}, \quad \mathcal{R} = \mathcal{C}_{\text{antigen}} \quad (7)$$

where c_{heavy} denotes the heavy chain and $\mathcal{C}_{\text{antigen}}$ represents the set of antigen chains.

5.3 Potential Interface Finder

To systematically explore the interaction landscape and resolve ambiguities in binding modes, we propose a Potential Interface Finder (PIF). This module analyzes the latent pairwise feature $\mathbf{Z}$ from an initial prediction to identify high-confidence binding sites, which are then converted into geometric constraints $\mathcal{K}$ to guide subsequent generative steps. The process consists of three phases: hotspot scoring, patch expansion, and hierarchical sampling.

Interaction Scoring and Similarity.

For a designated binder chain c_b and a candidate receptor region (set of tokens $\mathcal{R}$), we extract the corresponding sub-tensor from the pairwise representation $\mathbf{Z} \in \mathbb{R}^{N \times N \times d}$. Let $\mathbf{z}_{ij} \in \mathbb{R}^d$ denote the feature vector between a binder token $i \in c_b$ and a receptor token $j \in \mathcal{R}$. We define the interaction score s_j for each receptor token j as its maximum activation magnitude across all binder tokens, representing its likelihood of being an interface residue:

$$s_j = \max_{i \in c_b} \|\mathbf{z}_{ij}\|_{\infty} \quad (8)$$

where $\|\cdot\|_{\infty}$ denotes the maximum absolute value across the channel dimension. To identify continuous surface patches, we further compute the interaction profile similarity between any two receptor tokens $j, k \in \mathcal{R}$. Let $\mathbf{v}_j \in \mathbb{R}^{|c_b| \cdot d}$ be the flattened feature vector characterizing the interaction of receptor token j against the entire binder chain. The similarity metric S_{jk} is defined as the cosine similarity between their

507 profiles:

$$508 \quad S_{jk} = \frac{\mathbf{v}_j \cdot \mathbf{v}_k}{\|\mathbf{v}_j\| \|\mathbf{v}_k\|} \quad (9)$$

510 High S_{jk} implies that tokens j and k share a similar “view” of the binder, suggesting
511 they belong to the same binding patch.

512 **Patch Expansion and Clustering.**

513 We employ a seed-and-expand strategy to delineate distinct interface patches:

- 514 1. **Hotspot Identification:** We iteratively select the receptor token j^* with the
515 highest score s_{j^*} (among unassigned tokens) as a seed.
- 516 2. **Patch Expansion:** The seed aggregates neighboring tokens k that satisfy two
517 criteria: (1) high interaction profile similarity ($S_{j^*k} > \tau$) and (2) spatial proxim-
518 ity based on the predicted contact probability ($P_{20}[j^*, k] > \delta$). This results in a
519 candidate patch $\mathcal{I}$, which is a set of contact receptor token indices.
- 520 3. **Redundancy Filtering:** To ensure diversity, we perform a second-level clustering.
521 Overlapping patches or those with high spatial connectivity are merged, and the
522 representative patch with the highest cumulative score is retained.

523 **Constraint Generation and Sampling.**

524 The identified patches are converted into geometric constraints to guide struc-
525 ture generation. For the top- M ranked patches, we construct a set of constraints
526 $\{\mathcal{K}_1, \dots, \mathcal{K}_M\}$, where each $\mathcal{K}_m = (c_b, \mathcal{I}^{(m)})$. The final inference follows a hierarchical
527 sampling protocol:

- 529 • **Step 1 (Exploration):** For a given random seed ϵ , we generate an initial prediction
530 to compute $\mathbf{Z}_\epsilon$ and identify the top $M = 5$ PIs.
- 531 • **Step 2 (Refinement):** For each identified interface constraint $\mathcal{K}_m$, we generate
532 $K = 5$ structural samples using the constraint-aware workflow described in Eq. (6).

533 This strategy effectively decomposes the complex global optimization problem into
534 localized sub-problems, ensuring that the model explicitly explores the most probable
535 binding modes suggested by its own latent features.

536

537 **5.4 Interface Ranking Score**

538 Standard ranking scores often exhibit bias when explicit geometric constraints are
539 applied, as the model may hallucinate high confidence to satisfy the input conditions
540 regardless of biological plausibility. To mitigate this, we introduce the Interface Rank-
541 ing Score (IfaceRanking), a consensus-based metric that evaluates the reproducibility
542 of binding modes across stochastic runs.

543 **Clustering via Jaccard Distance.** We quantify the geometric dissimilarity
544 between two predicted interfaces, defined as contact token sets $\mathcal{I}_i$ and $\mathcal{I}_j$, using the
545 Jaccard distance:

$$546 \quad d(\mathcal{I}_i, \mathcal{I}_j) = 1 - \frac{|\mathcal{I}_i \cap \mathcal{I}_j|}{|\mathcal{I}_i \cup \mathcal{I}_j|} \quad (10)$$

547 An adaptive DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
548 algorithm is then applied to the pairwise distance matrix to identify dense clusters of
549 consistent binding interfaces, where the neighborhood radius ϵ is dynamically adjusted
550
551
552

to find stable groups. DBSCAN is an unsupervised machine learning algorithm that groups closely packed data points, defining clusters as high-density regions separated by low-density regions.

Hybrid Selection Strategy. The final model selection employs a hybrid approach to balance confidence and diversity. The Top-1 is identified simply as the sample with the highest intrinsic ranking score. Conversely, for the Top-5, we utilize the IfaceRanking clusters to filter for distinct binding modes. We first select one representative constraint from each identified cluster and subsequently rank these representatives by their intrinsic scores. The top-5 candidates are then chosen from this re-ranked list, ensuring the output ensemble captures structurally diverse high-confidence interfaces rather than redundant sampling of a single mode.

6 Data availability

The datasets generated and analyzed during the current study have been deposited in the Zenodo repository at [URL]. Protein structures used for training and evaluation were obtained from the RCSB Protein Data Bank (PDB) (<https://www.rcsb.org>). The specific PDB accession codes for the case studies discussed in the text are 7R58, 7UPM, and 7EDS. Source data are provided with this paper.

7 Code availability

All source code for the DynaHelix model, including training inference scripts, and data analysis notebooks, is available on GitHub at [GitHub URL]. A persistent release of the code used in this study is archived on Zenodo at [DOI]. Any additional custom scripts are available from the corresponding author upon reasonable request.

8 Figures

9 Acknowledgments

This work was supported by AI for Life Sciences Lab, Tencent, Shenzhen, China. This work was supported by the National Natural Science Foundation of China (62425204, U22A2037, 62450002, 62432011). We acknowledge each colleague who helped in this work.

10 Contributions

H.L. and X.Z. conceived the study. H.L. trained the model, and wrote the manuscript. H.L. and P.Z. designed the methodology, conducted model evaluations, and performed data analysis. H.J. and H.L. identified the core research problem. W.L., A.N., and X.Z. supervised the project. All authors, including K.X., and G.Y. reviewed the results and contributed to the critical revision of the manuscript.

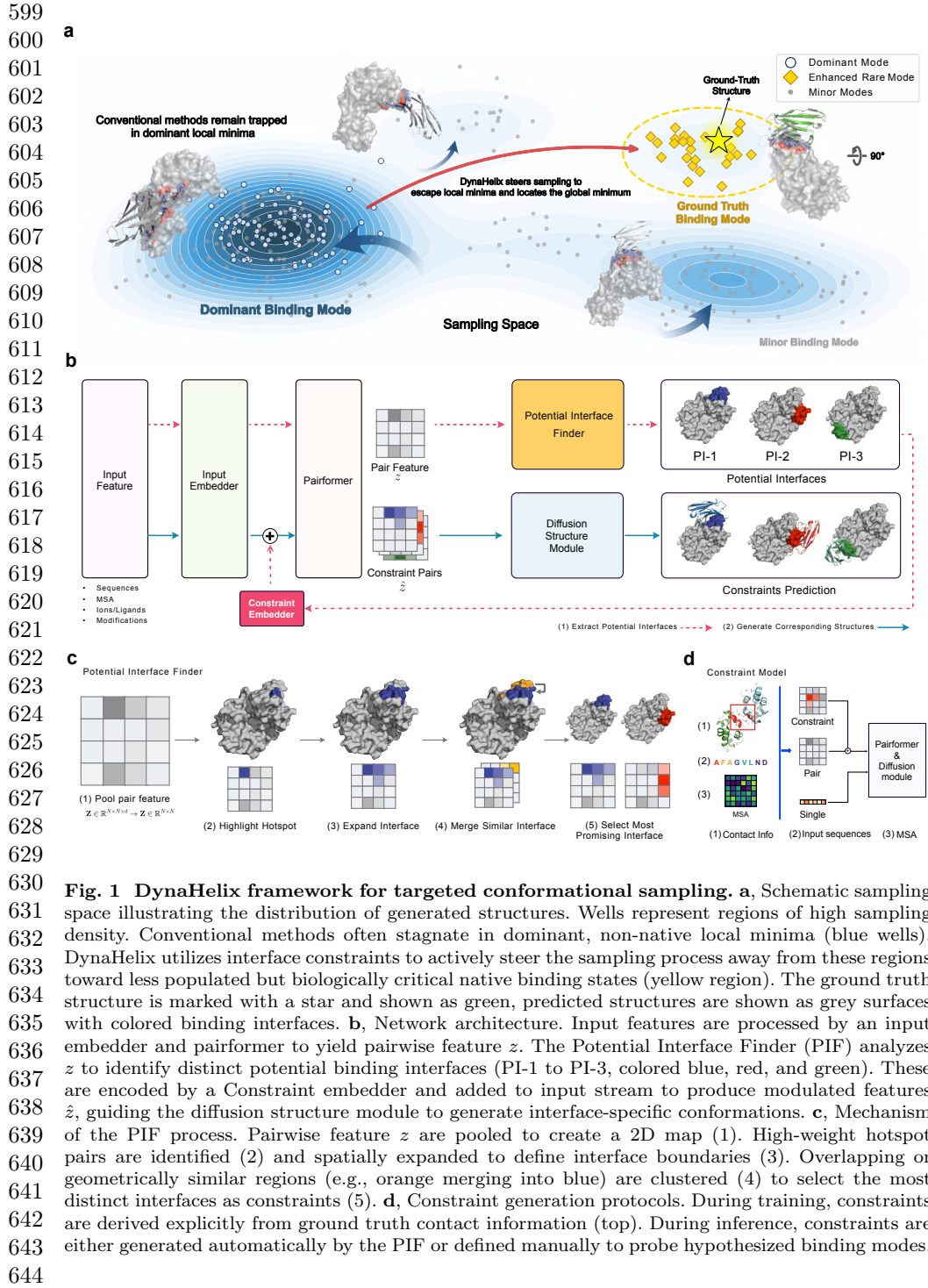


Fig. 1 DynaHelix framework for targeted conformational sampling. **a**, Schematic sampling space illustrating the distribution of generated structures. Wells represent regions of high sampling density. Conventional methods often stagnate in dominant, non-native local minima (blue wells). DynaHelix utilizes interface constraints to actively steer the sampling process away from these regions toward less populated but biologically critical native binding states (yellow region). The ground truth structure is marked with a star and shown as green, predicted structures are shown as grey surfaces with colored binding interfaces. **b**, Network architecture. Input features are processed by an input embedder and pairformer to yield pairwise feature z . The Potential Interface Finder (PIF) analyzes z to identify distinct potential binding interfaces (PI-1 to PI-3, colored blue, red, and green). These are encoded by a Constraint embedder and added to input stream to produce modulated features $\hat{z}$, guiding the diffusion structure module to generate interface-specific conformations. **c**, Mechanism of the PIF process. Pairwise feature z are pooled to create a 2D map (1). High-weight hotspot pairs are identified (2) and spatially expanded to define interface boundaries (3). Overlapping or geometrically similar regions (e.g., orange merging into blue) are clustered (4) to select the most distinct interfaces as constraints (5). **d**, Constraint generation protocols. During training, constraints are derived explicitly from ground truth contact information (top). During inference, constraints are either generated automatically by the PIF or defined manually to probe hypothesized binding modes.

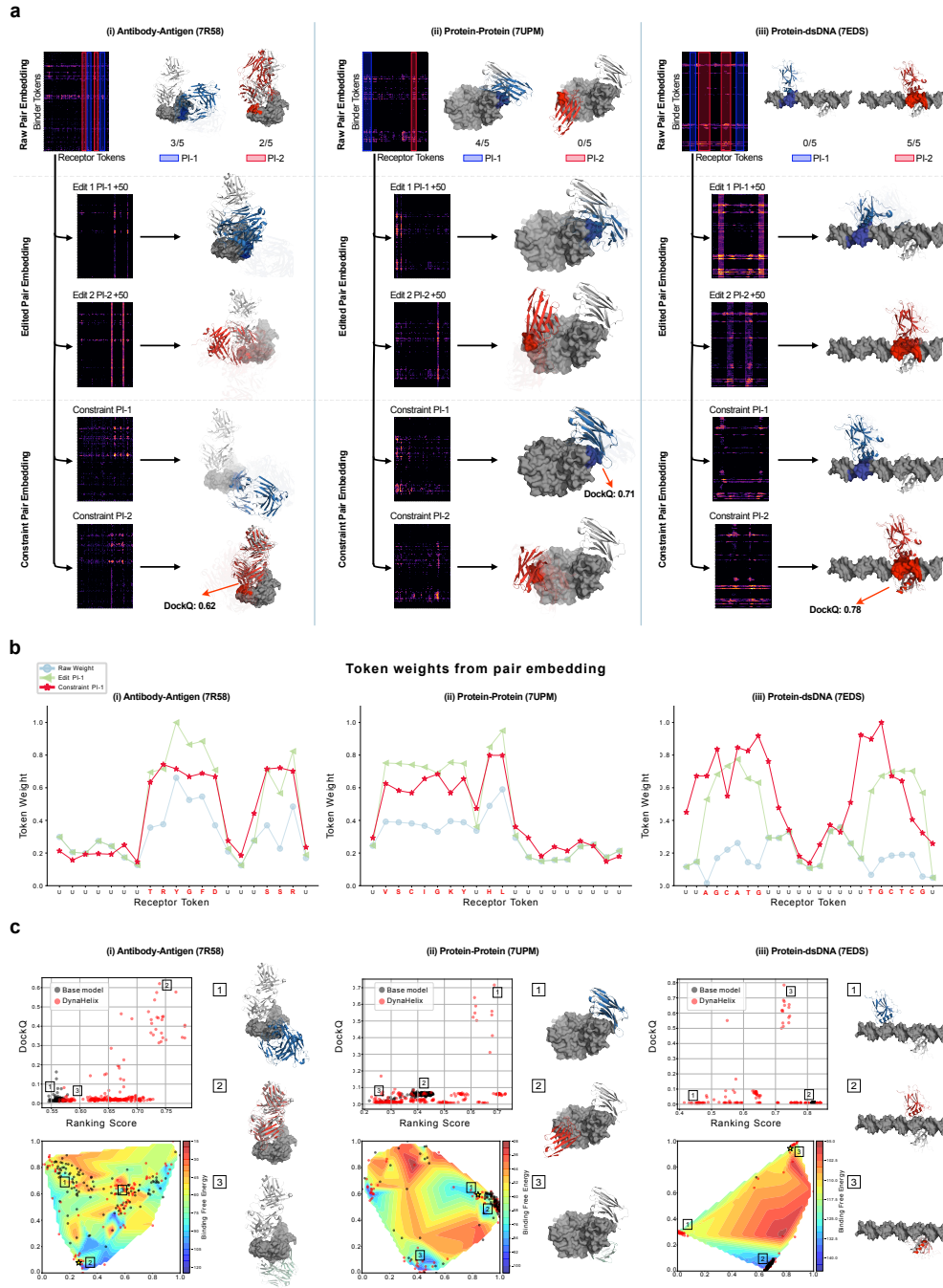


Fig. 2 Mechanism of latent feature modulation and enhanced sampling efficacy via DynaHelix. **a**, Visualization of the latent pairwise feature space z and resulting structural predictions across three interaction modalities: antibody-antigen (PDB 7R58), protein-protein (7UPM), and protein-dsDNA (7EDS). Rows compare raw features (baseline), manually perturbed weights (adding 50 bias to interface pairs), and constraint-guided modulation. Brighter regions in the heatmaps represent higher predicted contact probabilities and feature magnitudes. While manual editing distorts geometry, constraint-based modulation generates physically valid structures (high DockQ) that align with the ground truth (grey cartoon). The receptor is shown as a grey surface, predicted binders are colored blue (PI-1) and red (PI-2). **b**, Comparative profile of contact probabilities along the receptor sequences. The x-axis explicitly differentiates specific interface tokens (red labels) from average non-interacting flanking regions (aggregated as 'U', black labels). Unlike manual editing (green), which results in disjointed, artificial spikes, constraint embedding (red) induces a coherent, biologically plausible elevation of contact probability across the binding site. **c**, Evaluation of sampling landscapes and ranking capability. Left top: Scatter plots of structural accuracy (DockQ) versus model confidence (Ranking Score). DynaHelix (red points) successfully samples high-quality conformations in regions inaccessible to the base model (grey points). Left bottom: Energy landscapes computed via MM/GBSA and projected onto a principal component analysis (PCA) of pairwise distances. Right: Representative structures corresponding to distinct local minima (labels 1-3) identified in the energy landscape.

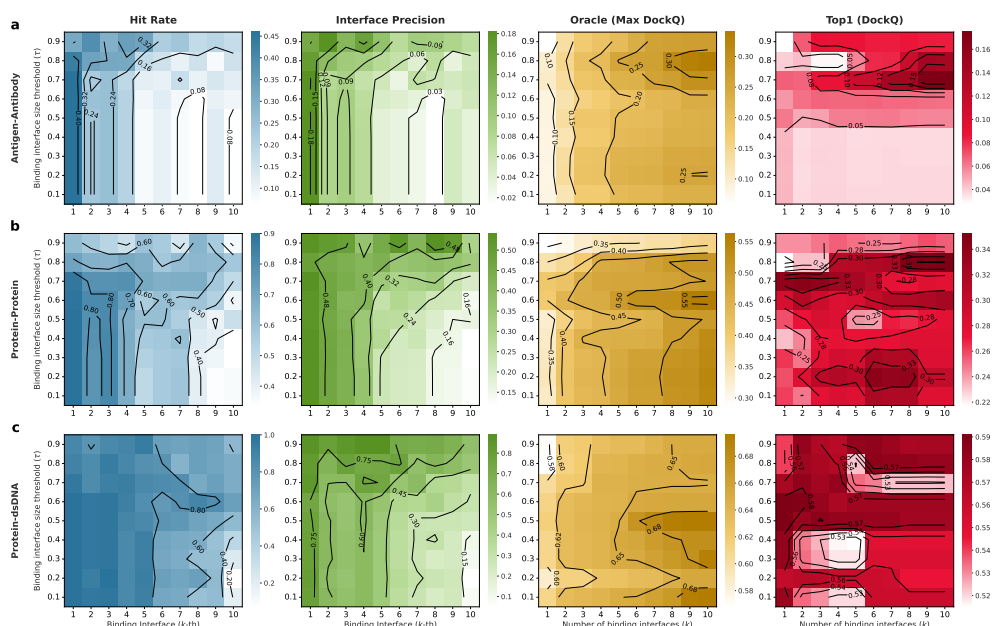


Fig. 3 Automated discovery of potential binding interfaces across interaction modalities. (a-c), Interface discovery performance heatmaps for antibody-antigen (a), protein-protein (b), and protein-dsDNA (c) interactions. The panels illustrate the retrieval efficacy as a function of the binding interface size threshold (τ , y -axis) and the number of sampling binding interfaces (k , x -axis). **Columns from left to right:** Hit Rate, the frequency of successfully identifying at least one ground-truth interface token; Interface Precision, measuring the signal purity of the recovered contact tokens; Oracle DockQ, the maximum DockQ score achieved among sampled candidates; and Top-1 DockQ, the accuracy of the highest-ranked prediction. Darker colors indicate higher performance values, and contour lines denote iso-performance levels, highlighting that the optimal discovery parameters are governed by the distinct physicochemical architectures of each interaction type.

11 Competing interests

The authors declare no competing interests.

Supplementary Information

Table of Contents

- [A](#) Benchmarks
- [B](#) Training protocol
- [C](#) Additional experiments

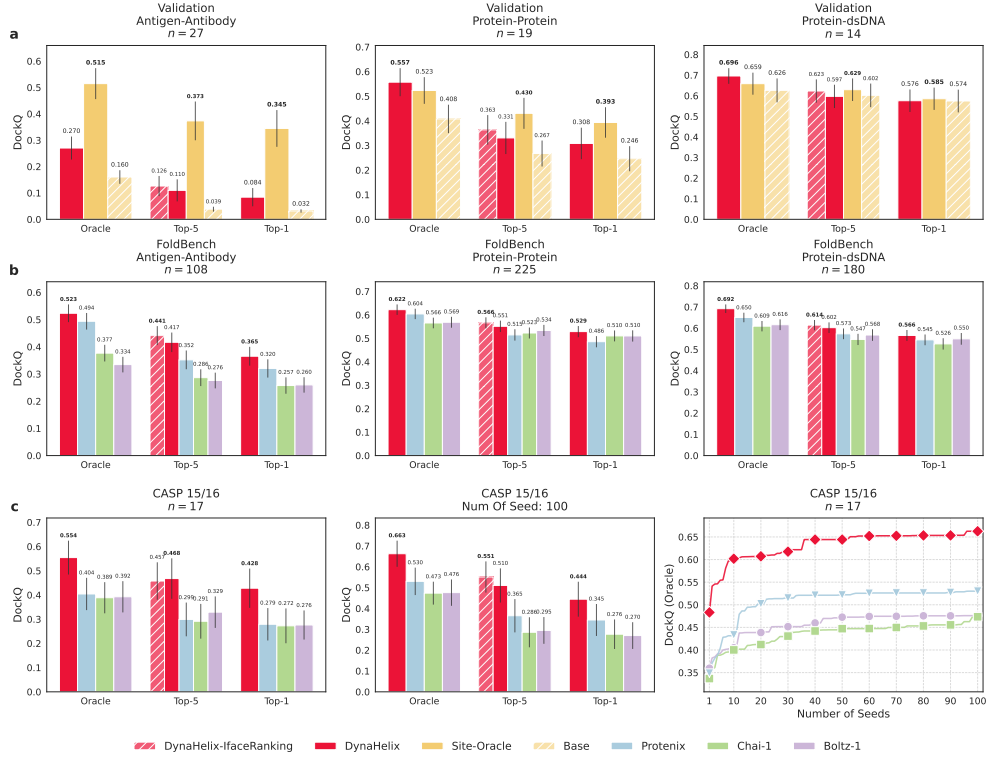


Fig. 4 Benchmarking DynaHelix against SOTA methods across diverse datasets. **a**, Performance on an internal low-homology validation set comprising antibody-antigen, protein-protein, and protein-dsDNA complexes. DynaHelix is compared with the base model and a reference using ground-truth interface constraints. **b**, Evaluation on the FoldBench benchmark comparing DynaHelix against leading models, including Protenix, Chai-1, and Boltz-1, across the same interaction categories. **c**, Assessment on CASP15 and CASP16 heteromultimer targets. The rightmost panel illustrates the cumulative maximum DockQ score (Oracle) as a function of the number of sampling seeds (1 to 100), demonstrating the sampling efficiency of DynaHelix relative to other methods. DockQ evaluation was restricted to antibody-antigen and protein-dsDNA pairs, whereas all pairwise interactions were assessed for protein-protein complexes. In all bar charts, data represent mean DockQ scores for Oracle, Top-5, and Top-1 ranked predictions; error bars denote the standard error of the mean (s.e.m.).

A Benchmarks

A.1 Validation

The validation set was derived from the dataset curated by Protenix, comprising 71 antibody-antigen, 325 protein-protein, and 38 dsDNA-protein structures. Driven by the hypothesis that the dominant predicted conformation does not necessarily correspond to the native ground truth, we deliberately targeted complexes that elude accurate prediction by current models. To strike a balance between computational efficiency and evaluative stringency, we systematically curated a highly representative

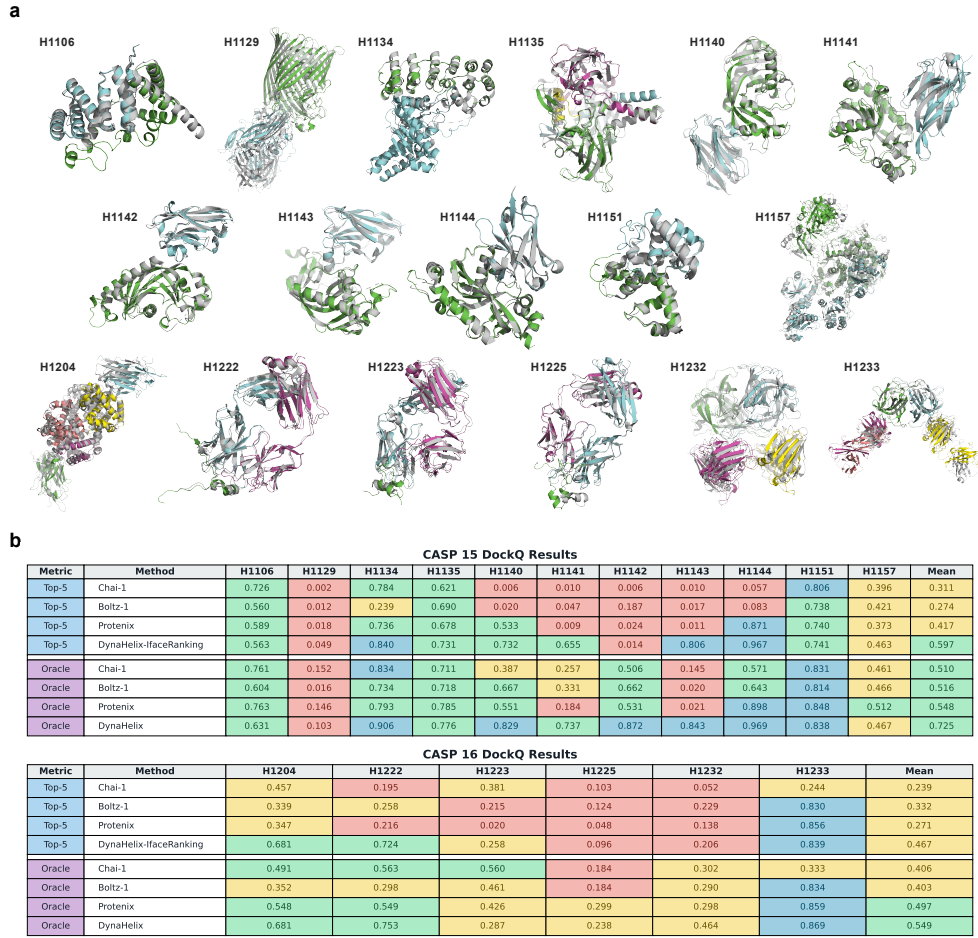


Fig. 5 DynaHelix performance on CASP15 and CASP16 heteromultimer targets. **a**, Superposition of DynaHelix predicted structures (colored chains) with corresponding ground-truth structures (grey) for selected CASP15 and CASP16 targets. **b**, Comprehensive DockQ score tables comparing DynaHelix against SOTA methods (Chai-1, Boltz-1, Protenix) across individual targets. Performance is evaluated using the best structure from the top-5 ranked predictions (Top-5) and the best structure among all generated samples (Oracle). All models utilized 100 sampling seeds. Table cells are color-coded based on DockQ score quality: High quality (>0.8 , blue), Medium quality ($0.49-0.8$, green), Low quality ($0.23-0.49$, yellow), and Incorrect/Failed (<0.23 , red). The mean score for each method is provided in the rightmost column.

subset of these structurally challenging targets. This subset, utilized for robust model validation and comprehensive parameter analysis, was selected based on the following criteria:

- **Antibody-Antigen:** 27 complexes were selected based on 25 models generated by Protenix, filtering for an DockQ < 0.23 and a total token count $< 1,000$.

- **Protein-Protein:** The 20 lowest-scoring complexes were selected based on LDDT rankings derived from 125 models generated by Protenix.
- **dsDNA-Protein:** 14 complexes were selected based on 125 models generated by Protenix, filtering for an DockQ < 0.8 and a total token count < 1,000.

A.2 FoldBench

The FoldBench dataset initially included 113 antibody-antigen, 239 protein-protein, and 190 dsDNA-protein complexes. Due to prediction failures in high-throughput experiments across certain models, we restricted our reporting to targets where all models generated valid outputs. Consequently, the final reported dataset comprises 108 antibody-antigen, 225 protein-protein, and 180 dsDNA-protein complexes; details of the failed PDBs are provided in Table S1.

Table S1 PDBs with failed predictions by different models on FoldBench.

	FB_AA	FB_PP	FB_dsDNA
Chai-1	8oud, 8udz, 8pn0, 8j7t, 8xnh	8b46, 8ic7, 9fvt, 8ff9, 8oq1, 8rja, 8c0h, 8x3g	8sv3, 8oj7, 8oj6, 8g88, 8iyq, 8oja, 8wt6, 9enp
Boltz-1	8oud, 8udz, 8pn0, 8j7t, 8xnh	8b46, 8ic7, 9fvt, 8ff9, 8oq1, 8rja, 7x36, 8c0h, 8a8o, 8x3g, 8iq5	8sv3, 8oj7, 8oj6, 8g88, 8oja, 8wt6, 9enp
Protenix	-	8qab, 8r7i	9cih, 8vdh
DynaHelix	-	8qab, 8gqp, 8r7i	9cih

A.3 CASP

We selected multimers from the official CASP15 and CASP16 repositories to serve as an additional test set. The inclusion criteria were as follows: (1) availability of public PDB reference structures; (2) a maximum of six chains to ensure high-throughput feasibility; and (3) an unresolved residue ratio of less than 60% per reference chain. This resulted in the selection of 11 complexes from CASP15 and 6 from CASP16.

The specific targets and their corresponding PDB IDs are as follows: H1106 (7qih), H1129 (8a8c), H1134 (7ubz), H1135 (7z8y), H1140 (9ert), H1141 (9eru), H1142 (9erw), H1143 (9etj), H1144 (9etl), H1151 (8d5v), H1157 (8pko), H1204 (8vyl), H1222 (9cqd), H1223 (9cqb), H1225 (9cqa), H1232 (9cn2), H1233 (9cbn).

A.4 Inference protocol

To ensure a rigorous comparison, we standardized the sampling budget to 125 candidate structures per target across all models. For DynaHelix, this ensemble was constructed by combining 5 random seeds with the top-5 predicted interfaces

(5 seeds $\times$ 5 PIs $\times$ 5 samples). For the baseline models (Protenix, Chai-1, and Boltz-1), the budget was allocated as 5 seeds $\times$ 25 samples. A consistent recycling depth of $N_{\text{cycles}} = 10$ was applied uniformly across all experiments.

To ensure consistency, all models utilized the ColabFold API [47] for MSA generation.

Inference efficiency was optimized following the Protenix-Mini protocol [48]. For both DynaHelix and Protenix, the diffusion sampling trajectory was condensed from 200 to 10 ODE steps, with noise parameters set to $\gamma = 0.0$ and $\text{step_scale_eta} = 1.0$. This configuration has been shown to significantly reduce computational cost with negligible performance impact. In contrast, Chai-1 and Boltz-1 were executed using their standard default parameters.

Predictions were evaluated using PXMeter [49], which calculates scores based on DockQ v2 [50]. To better reflect real-world applicability, we report the mean DockQ across all predicted conformers for each complex instead of averaging the scores at the interface level. For antibody-antigen complexes in the validation and CASP datasets, DockQ scores were calculated exclusively between the antigen and the antibody heavy or light chains. For the protein-dsDNA validation set, we computed DockQ between the protein and each individual DNA strand. Finally, for the FoldBench dataset, although evaluations targeted the specific interfaces defined by the benchmark, the final metric remains rigorously averaged per complex.

For Antibody-Antigen and Protein-Protein datasets, we employed the official Protenix weights ([protenix_base_constraint_v0.5.0](#)); for Protein-dsDNA datasets, the model was retrained (details in section B) to address the official weights’ inability to handle DNA constraints. Versions of all open-source software are listed in Table S2.

Table S2 GitHub Repository and Commit Hash of Evaluated Models

Model	GitHub Repository	Commit Hash
Protenix	bytedance/Protenix	fe26d544
Chai-1	chaidiscovery/chai-lab	c940e9df
Boltz-1	jwohlwend/boltz	397fa2ac
PXMeter	bytedance/PXMeter	80cae617

B Training protocol

B.1 Constraint models training

Our model was built upon the Protenix framework (version 0.5.0, commit [fe26d544](#)) and extended to incorporate constraint-based features compatible with various molecular types, including proteins and dsDNA.

For training, we utilized the [weightedPDB_before2109_wopb_nometalc_0925](#) dataset provided by Protenix. This dataset comprises all Protein Data Bank (PDB) entries released prior to September 30, 2021, maintaining a data cutoff consistent with

the configuration of AlphaFold 3. To validate the model’s performance, we employed the `recentPDB.1536_sample384.0925` dataset.

A key component of our training strategy involved the probabilistic injection of constraints to enable controlled structure generation. During the training phase, constraints were applied with the following probabilities:

- **10% probability:** Atom-atom constraints extracted directly from PDB structures.
- **20% probability:** Binder chain and receptor token constraints, as described in the methodology section.
- **10% probability:** Token-token interaction constraints.

The training process was conducted over a period of approximately two weeks using a cluster of 32 NVIDIA H20 GPUs.

B.2 Hyperparameter configuration

The model was trained using the BF16 mixed-precision format to optimize computational efficiency. We employed an Exponential Moving Average (EMA) with a decay rate of 0.999 to stabilize the training process. The learning rate was initialized at 1×10^{-3} with a warmup period of 2,000 steps, followed by a step-based decay every 20,000 steps. The detailed hyperparameters used in our experiments are summarized in Table S3.

Table S3 Hyperparameter settings for model training and sampling.

Hyperparameter	Value	Description
Precision Type	BF16	Training data type format
Diffusion Batch Size	48	Batch size used for the diffusion module
EMA Decay	0.999	Exponential moving average decay rate
Crop Size	384	Training crop size for input sequences
Max Steps	100,000	Total number of training steps
Warmup Steps	2,000	Number of steps for learning rate warmup
Learning Rate (LR)	0.001	Initial base learning rate
LR Decay Interval	20,000	Step interval for learning rate decay
Sampling Steps	20	Number of diffusion steps

B.3 Sampling Hot interfaces as decoys to enhance confident head

The introduction of constraint features may adversely affect the confidence head, as alternative interface constraints are rarely encountered during standard training and may therefore be treated as outliers. To address this, we incorporate the PIF algorithm with a 20% activation probability during each training step. When activated, an interface constraint is randomly selected, prompting the structure module to generate a corresponding conformation based on that specific constraint. This strategy exposes the confidence head to a broader sampling space beyond the native PDB

structures, thereby enhancing its generalization capabilities. As experiments regarding this approach are currently in progress, we plan to release these updated models in the future.

B.4 MM/GBSA binding free energy calculation

The binding free energies of the complexes were estimated using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method with the OpenMM (version 8.2.0) [51]. To describe atomic interactions, the ff14SB force field [52] was applied to proteins, and the OL15 force field [53] was used for DNA. The solvent effects were treated using the GBn2 implicit solvent model [54].

Prior to the energy calculation, each complex underwent an energy minimization process (L-BFGS, up to 5000 iterations) in the presence of the implicit solvent to resolve steric clashes and relax the local geometry of the predicted structures. The binding free energy (ΔG_{bind}) was calculated based on the minimized structures using the following equation:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{ligand}}) \quad (11)$$

where G is the total potential energy calculated using the specified force fields and the GBn2/SA model. Receptor and ligand coordinates were extracted from the minimized complex (single-trajectory), and conformational entropy was not included. For each complex structure, the minimization and ΔG_{bind} estimation were repeated in five replicas to ensure numerical stability.

C Additional Experiments

C.1 Examples of better binding interface sampling by DynaHelix

Figure S1 illustrates representative sampling outcomes across various models. For certain complexes (e.g., 7YF2, 8U0O, 9C3T, 8URK, 9C3S, 8UP2), baseline models failed to capture the correct binding interface ($\text{DockQ} < 0.23$), whereas DynaHelix successfully generated valid solutions. In instances where baselines produced only low-quality interfaces ($0.23 < \text{DockQ} < 0.49$), DynaHelix improved results to medium quality or above (e.g., 8ET0, 8FYM, 8T59, 8EKZ, 8IDO, 8HXQ, 7TXW, 8H0L). Furthermore, for targets where baselines achieved medium quality ($0.49 < \text{DockQ} < 0.8$), DynaHelix further refined these into high-quality conformations (e.g., 8AON, 8J7T).

C.2 DynaHelix outperforms baseline models across different backbones

DynaHelix extends readily to other constraint-compatible, AlphaFold3-like architectures, such as Chai-1. We conducted a comprehensive ablation study across various benchmarks to evaluate performance with and without DynaHelix. As illustrated in Fig. S2, integrating DynaHelix with Protenix yielded consistent improvements in

1013 Oracle, Top-5, and Top-1 metrics on FoldBench. Notably, the most significant gains
1014 occurred in the antibody-antigen subset, where mean DockQ scores increased by 0.07
1015 (Top-5) and 0.052 (Top-1).

1016 These improvements were even more pronounced when using Chai-1 as the back-
1017 bone. On the antibody-antigen dataset, we observed mean increases of 0.118 (Oracle),
1018 0.09 (Top-5), and 0.091 (Top-1). Significant gains were also evident on the CASP15/16
1019 datasets, where DynaHelix consistently outperformed baseline models as the number
1020 of sampling seeds increased.

1021

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

1057

1058

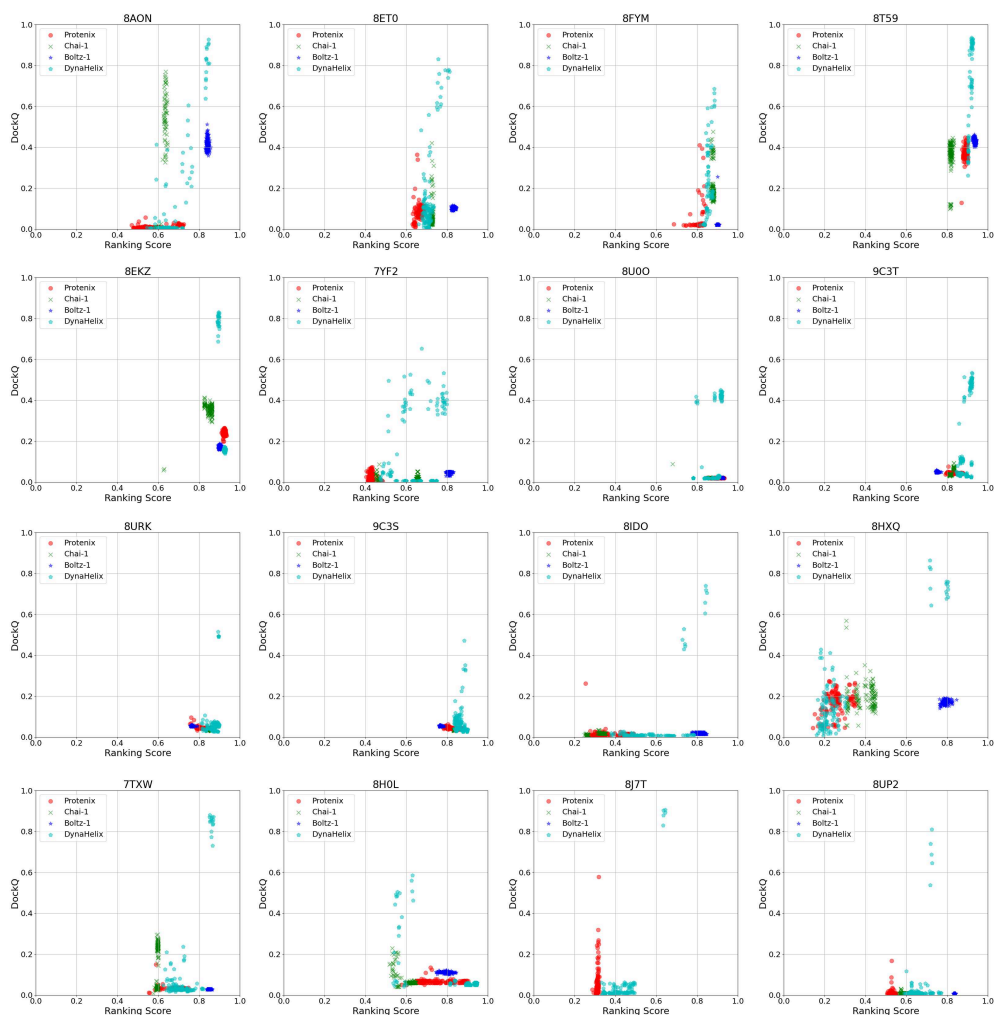


Fig. S1 Scatter plot of DockQ versus Ranking Score for the 16 test samples. DynaHelix exhibits a sampling distribution structured into distinct clusters, notably encompassing conformations close to the native PDB binding mode. In contrast, baseline models display a highly concentrated distribution, suggesting a susceptibility to bias that hinders their ability to locate the correct binding interface.

References

- [1] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstern, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D.: Highly accurate protein

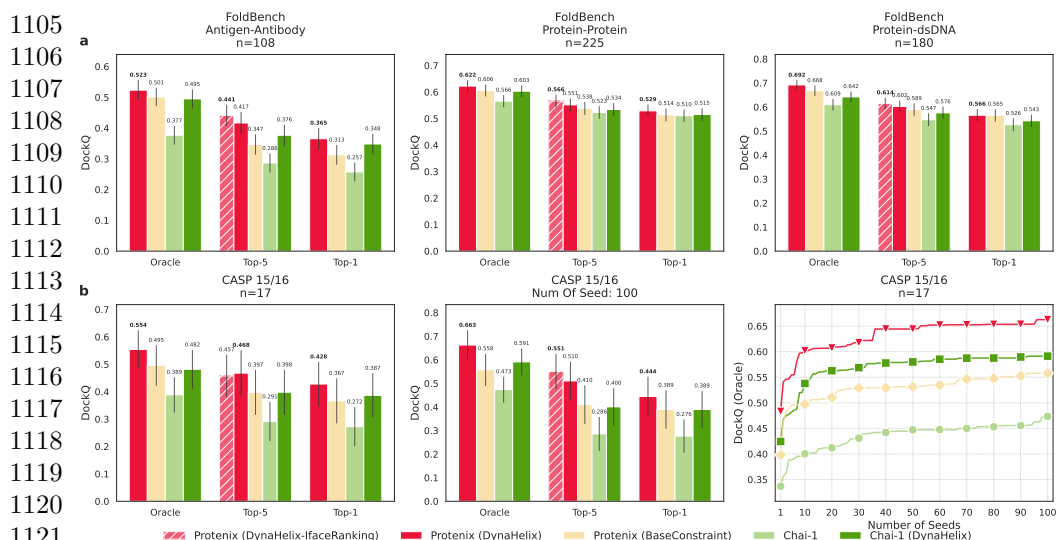


Fig. S2 Performance of DynaHelix with Different Backbones. DynaHelix delivers consistent and substantial improvements across various datasets, regardless of whether Protenix or Chai-1 is employed as the backbone model.

- structure prediction with AlphaFold. Nature **596**(7873), 583–589 (2021)
- [2] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., Hassabis, D.: Protein complex prediction with AlphaFold-Multimer (2021)
- [3] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Židek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M.: Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature **630**(8016), 493–500 (2024)
- [4] Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G.R., Morey-Burrows, F.S., Anishchenko, I., Humphreys, I.R., McHugh, R., Vafeados, D., Li, X., Sutherland, G.A., Hitchcock, A., Hunter, C.N., Kang, A., Brackenbrough, E., Bera, A.K., Baek, M., DiMaio, F., Baker, D.: Generalized biomolecular modeling and design with RoseTTAFold All-Atom. Science

384(6693), 2528 (2024)	1151
	1152
[5] Roney, J.P., Ovchinnikov, S.: State-of-the-art estimation of protein model accuracy using AlphaFold. <i>Phys. Rev. Lett.</i> 129 (23), 238101 (2022)	1153
	1154
	1155
[6] Kryshchak, A., Schwede, T., Topf, M., Fidelis, K., Moult, J.: Critical assessment of methods of protein structure prediction (CASP)-Round XV. <i>Proteins</i> 91 (12), 1539–1549 (2023)	1156
	1157
	1158
	1159
[7] Wallner, B.: AFsample: improving multimer prediction with AlphaFold using massive sampling. <i>Bioinformatics</i> 39 (9) (2023)	1160
	1161
	1162
[8] Wallner, B.: Improved multimer prediction using massive sampling with AlphaFold in CASP15. <i>Proteins</i> 91 (12), 1734–1746 (2023)	1163
	1164
	1165
[9] Basu, S., Wallner, B.: DockQ: A quality measure for protein-protein docking models. <i>PLoS One</i> 11 (8), 0161879 (2016)	1166
	1167
	1168
[10] Yuan, R., Zhang, J., Kryshchak, A., Schaeffer, R.D., Zhou, J., Cong, Q., Grishin, N.V.: CASP16 protein monomer structure prediction assessment. <i>bioRxiv</i> (2025)	1169
	1170
	1171
[11] Alexander, L.T., Follonier, O.M., Kryshchak, A., Abesamis, K., Bibi-Triki, S., Box, H.G., Breyton, C., Bringel, F., Carrique, L., d’Acapito, A., Dong, G., DuBois, R., Fass, D., Fiesco, J.M., Fox, D.R., Grimes, J.M., Grinter, R., Jenkins, M., Kamyschinsky, R., Keown, J.R., Lackner, G., Lammers, M., Liu, S., Lovering, A.L., Malinauskas, T., Masquida, B., Palm, G.J., Siebold, C., Su, T., Zhang, P., Zhou, Z.H., Fidelis, K., Topf, M., Moult, J., Schwede, T.: Protein target highlights in CASP16: Insights from the structure providers. <i>Proteins</i> 94 (1), 25–50 (2026)	1172
	1173
	1174
	1175
	1176
	1177
	1178
	1179
[12] Zhang, J., Yuan, R., Kryshchak, A., Kretsch, R.C., Schaeffer, R.D., Zhou, J., Das, R., Grishin, N.V., Cong, Q.: Assessment of Protein Complex Predictions in CASP16: Are we making progress? (2025)	1180
	1181
	1182
	1183
[13] Chai Discovery: Chai-1: Decoding the molecular interactions of life. <i>bioRxiv</i> (2024) https://doi.org/10.1101/2024.10.10.615955	1184
	1185
	1186
[14] Wohlwend, J., Corso, G., Passaro, S., Getz, N., Reveiz, M., Leidal, K., Swiderski, W., Atkinson, L., Portnoi, T., Chinn, I., Silterra, J., Jaakkola, T., Barzilay, R.: Boltz-1: Democratizing biomolecular interaction modeling. <i>bioRxiv</i> (2024) https://doi.org/10.1101/2024.11.19.624167	1187
	1188
	1189
	1190
	1191
[15] Team, B.A.A., Chen, X., Zhang, Y., Lu, C., Ma, W., Guan, J., Gong, C., Yang, J., Zhang, H., Zhang, K., Wu, S., Zhou, K., Yang, Y., Liu, Z., Wang, L., Shi, B., Shi, S., Xiao, W.: Protenix - advancing structure prediction through a comprehensive alphafold3 reproduction. <i>bioRxiv</i> (2025) https://doi.org/10.1101/2025.01.08.631967	1192
	1193
	1194
	1195
	1196

- 1197 [16] Kalakoti, Y., Wallner, B.: AFsample2 predicts multiple conformations and
1198 ensembles with AlphaFold2. *Commun. Biol.* **8**(1), 373 (2025)
1199
- 1200 [17] Raouraoua, N., Mirabello, C., Véry, T., Blanchet, C., Wallner, B., Lensink,
1201 M.F., Brysbaert, G.: MassiveFold: unveiling AlphaFold’s hidden potential with
1202 optimized and parallelized massive sampling. *Nat. Comput. Sci.* **4**(11), 824–828
1203 (2024)
1204
- 1205 [18] Del Alamo, D., Sala, D., Mchaourab, H.S., Meiler, J.: Sampling alternative confor-
1206 mational states of transporters and receptors with AlphaFold2. *Elife* **11**(e75751)
1207 (2022)
- 1208 [19] Silva, G., Cui, J.Y., Dalgarno, D.C., Lisi, G.P., Rubenstein, B.M.: High-
1209 throughput prediction of protein conformational distributions with subsampled
1210 AlphaFold2. *Nat. Commun.* **15**(1), 2464 (2024)
1211
- 1212 [20] Stein, R.A., Mchaourab, H.S.: SPEACH_AF: Sampling protein ensembles and
1213 conformational heterogeneity with alphafold2. *PLoS Comput. Biol.* **18**(8),
1214 1010483 (2022)
1215
- 1216 [21] Wayment-Steele, H.K., Ojoawo, A., Otten, R., Apitz, J.M., Pitsawong, W.,
1217 Hömberger, M., Ovchinnikov, S., Colwell, L., Kern, D.: Predicting multiple con-
1218 formations via sequence clustering and AlphaFold2. *Nature* **625**(7996), 832–839
1219 (2024)
1220
- 1221 [22] Tao, E., Corry, B.: Alphafold2 captures conformational transitions in the voltage-
1222 gated channel superfamily. *bioRxiv*, 2025–03 (2025)
1223
- 1224 [23] Stein, R.A., Mchaourab, H.S.: Speech_af: Sampling protein ensembles and con-
1225 formational heterogeneity with alphafold2. *PLOS Computational Biology* **18**(8),
1226 1–16 (2022) <https://doi.org/10.1371/journal.pcbi.1010483>
1227
- 1228 [24] Bryant, P., Noé, F.: Improved protein complex prediction with AlphaFold-
1229 multimer by denoising the MSA profile. *PLoS Comput. Biol.* **20**(7), 1012253
1230 (2024)
- 1231 [25] Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorney, D., Yueh, C., Beglov,
1232 D., Vajda, S.: The ClusPro web server for protein–protein docking. *Nat. Protoc.*
1233 **12**(2), 255–278 (2017)
1234
- 1235 [26] Honorato, R.V., Trellet, M.E., Jiménez-García, B., Schaarschmidt, J.J., Giulini,
1236 M., Reys, V., Koukos, P.I., Rodrigues, J.P.G.L.M., Karaca, E., Zundert, G.C.P.,
1237 Roel-Touris, J., Noort, C.W., Jandová, Z., Melquiond, A.S.J., Bonvin, A.M.J.J.:
1238 The HADDOCK2.4 web server for integrative modeling of biomolecular com-
1239 plexes. *Nat. Protoc.* **19**(11), 3219–3241 (2024)
1240
- 1241 [27] Lyskov, S., Gray, J.J.: The RosettaDock server for local protein-protein docking.
1242

Nucleic Acids Res. 36 (Web Server issue), 233–8 (2008)	1243
	1244
[28] Brigham, E.O., Morrow, R.E.: The fast fourier transform. <i>IEEE Spectr.</i> 4 (12), 63–70 (1967)	1245
	1246
	1247
[29] Vecchietti, L.F., Wijaya, B.N., Armanuly, A., Hangeldiyev, B., Jung, H., Lee, S., Cha, M., Kim, H.M.: Artificial intelligence-driven computational methods for antibody design and optimization. <i>MAbs</i> 17 (1), 2528902 (2025)	1248
	1249
	1250
	1251
[30] Zhu, W., Shenoy, A., Kundrotas, P., Elofsson, A.: Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes. <i>Bioinformatics</i> 39 (7) (2023)	1252
	1253
	1254
	1255
[31] Mitra, R., Li, J., Sagendorf, J.M., Jiang, Y., Cohen, A.S., Chiu, T.-P., Glasscock, C.J., Rohs, R.: Geometric deep learning of protein-DNA binding specificity. <i>Nat. Methods</i> 21 (9), 1674–1683 (2024)	1256
	1257
	1258
	1259
[32] Hayes, T.L., Krishnan, G.P.: Quantifying the role of openfold components in protein structure prediction. In: <i>NeurIPS 2025 2nd Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences</i>	1260
	1261
	1262
[33] Wu, D., Feng, L.: Robust prediction of multiple protein conformations with entropy guidance. <i>bioRxiv</i> , 2025–04 (2025)	1263
	1264
	1265
[34] Suzuki, S., Amagasa, T.: Steering conformational sampling in boltz-2 via pair representation scaling. <i>bioRxiv</i> , 2026–01 (2026)	1266
	1267
	1268
[35] Billiald, P., Slater, A., Welin, M., Clark, J.C., Loyau, S., Pugnière, M., Jiacomini, I.G., Rose, N., Lebozec, K., Toledano, E., François, D., Watson, S.P., Jandrot-Perrus, M.: Targeting platelet GPVI with glenzocimab: a novel mechanism for inhibition. <i>Blood Adv.</i> 7 (7), 1258–1268 (2023)	1269
	1270
	1271
	1272
	1273
[36] Jamieson, S.A., Pudjihartono, M., Horne, C.R., Vilorio, J.S., Dunlop, J.L., McMillan, H.D., Day, R.C., Keeshan, K., Murphy, J.M., Mace, P.D.: Nanobodies identify an activated state of the TRIB2 pseudokinase. <i>Structure</i> 30 (11), 1518–15295 (2022)	1274
	1275
	1276
	1277
	1278
[37] Xing, Y.F., Lu, M.: Human p53 core domain with germline hot spot mutation M133T in complex with the natural PIG3 p53-response element and Arsenic. <i>Worldwide Protein Data Bank</i> (2022)	1279
	1280
	1281
	1282
[38] Sundberg, E.J.: Structural basis of antibody–antigen interactions. In: <i>Epitope Mapping Protocols. Methods in Molecular Biology</i> (Clifton, N.J.), pp. 23–36. Humana Press, Totowa, NJ (2009)	1283
	1284
	1285
	1286
[39] Sela-Culang, I., Kunik, V., Ofran, Y.: The structural basis of antibody-antigen recognition. <i>Front. Immunol.</i> 4 , 302 (2013)	1287
	1288

- 1289 [40] Chothia, C., Janin, J.: Principles of protein-protein recognition. *Nature*
1290 **256**(5520), 705–708 (1975)
1291
- 1292 [41] Prasad Bahadur, R., Chakrabarti, P., Rodier, F., Janin, J.: A dissection of specific
1293 and non-specific protein–protein interfaces. *J. Mol. Biol.* **336**(4), 943–955 (2004)
1294
- 1295 [42] Luscombe, N.M., Laskowski, R.A., Thornton, J.M.: Amino acid-base interac-
1296 tions: a three-dimensional analysis of protein-DNA interactions at an atomic level.
1297 *Nucleic Acids Res.* **29**(13), 2860–2874 (2001)
- 1298 [43] Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., Mann, R.S.: Origins of
1299 specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**(1), 233–269
1300 (2010)
1301
- 1302 [44] Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J.Z.H., Hou, T.: End-point
1303 binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and
1304 applications in drug design. *Chem. Rev.* **119**(16), 9478–9508 (2019)
1305
- 1306 [45] Eastman, P., Galvelis, R., Peláez, R.P., Abreu, C.R.A., Farr, S.E., Gallicchio, E.,
1307 Gorenko, A., Henry, M.M., Hu, F., Huang, J., Krämer, A., Michel, J., Mitchell,
1308 J.A., Pande, V.S., Rodrigues, J.P., Rodriguez-Guerra, J., Simmonett, A.C., Singh,
1309 S., Swails, J., Turner, P., Wang, Y., Zhang, I., Chodera, J.D., De Fabritiis,
1310 G., Markland, T.E.: OpenMM 8: Molecular dynamics simulation with machine
1311 learning potentials. *J. Phys. Chem. B* **128**(1), 109–116 (2024)
1312
- 1313 [46] Xu, S., Feng, Q., Qiao, L., Wu, H., Shen, T., Cheng, Y., Zheng, S., Sun, S.:
1314 Benchmarking all-atom biomolecular structure prediction with FoldBench. *Nat.*
1315 *Commun.* **17**(1), 442 (2025)
1316
- 1317 [47] Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger,
1318 M.: Colabfold: making protein folding accessible to all. *Nature methods* **19**(6),
1319 679–682 (2022)
1320
- 1321 [48] Gong, C., Chen, X., Zhang, Y., Song, Y., Zhou, H., Xiao, W.: Protenix-mini:
1322 Efficient structure predictor via compact architecture, few-step diffusion and
1323 switchable plm. *arXiv preprint arXiv:2507.11839* (2025)
- 1324 [49] Ma, W., Liu, Z., Yang, J., Lu, C., Zhang, H., Xiao, W.: From dataset curation
1325 to unified evaluation: Revisiting structure prediction benchmarks with pxmeter.
1326 *bioRxiv* (2025) <https://doi.org/10.1101/2025.07.17.664878>
1327
- 1328 [50] Mirabetto, C., Wallner, B.: Dockq v2: improved automatic quality measure for
1329 protein multimers, nucleic acids, and small molecules. *Bioinformatics* **40**(10), 586
1330 (2024) <https://doi.org/10.1093/bioinformatics/btae586>
1331
- 1332 [51] Eastman, P., Galvelis, R., Peláez, R.P., Abreu, C.R., Farr, S.E., Gallicchio, E.,
1333 Gorenko, A., Henry, M.M., Hu, F., Huang, J., *et al.*: Openmm 8: molecular
1334

dynamics simulation with machine learning potentials. The Journal of Physical Chemistry B 128 (1), 109–116 (2023)	1335 1336 1337
[52] Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., Simmerling, C.: ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. Journal of chemical theory and computation 11 (8), 3696–3713 (2015)	1338 1339 1340 1341 1342
[53] Zgarbová, M., Sponer, J., Otyepka, M., Cheatham III, T.E., Galindo-Murillo, R., Jurecka, P.: Refinement of the sugar–phosphate backbone torsion beta for amber force fields improves the description of z-and b-dna. Journal of chemical theory and computation 11 (12), 5723–5736 (2015)	1343 1344 1345 1346 1347
[54] Nguyen, H., Roe, D.R., Simmerling, C.: Improved generalized born solvent model parameters for protein simulations. Journal of chemical theory and computation 9 (4), 2020–2034 (2013)	1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380