Reshaping Biomolecular Structure Prediction through Strategic Conformational Exploration with HelixFold-S1

Lihang Liu[1†], Yang Liu[1†], Xianbin Ye[1†], Shanzhuo Zhang[1], Yuxin Li[1], Kunrui Zhu[1], Yang Xue[1], Jingbo Zhou[1], Xiaonan Zhang[1], Xiaomin Fang[1*]

[1]PaddleHelix Team, Baidu Inc., Shenzhen 518000, China.

*Corresponding author(s). E-mail(s): fangxiaomin666@gmail.com;
†These authors contributed equally to this work.

Abstract

Generating large ensembles of candidate conformations is standard for improving biomolecular structure prediction. Yet aimless sampling is inefficient and costly, producing many redundant conformations with limited diversity, particularly for complex multimeric assemblies. Here, we present HelixFold-S1, a guided planning approach specifically designed to enhance the structural prediction of biomolecular complexes by strategically targeting the most informative regions of conformational space to produce accurate conformations. For each complex, predicted inter-chain contact probabilities serve as a blueprint of the conformational space, guiding computational effort toward higher-probability, low-redundancy contacts that constrain structure generation. Across diverse biomolecular complex benchmarks, HelixFold-S1 achieves markedly higher structural accuracy than traditional unguided methods while reducing sampling requirements by an order of magnitude. Predicted contact probabilities also provide a rough indicator of prediction difficulty and sampling utility. These results demonstrate that guided planning reshapes conformational exploration and enables more efficient and accurate structural inference.

Keywords: Biomolecular structure prediction, Conformation space exploration, Planning-driven, Computational efficiency, Resource-aware modeling

Introduction

Biomolecular structure prediction lies at the heart of computational biology, enabling progress in drug discovery, protein engineering, and the study of molecular interactions. Deep learning has transformed this field, with recent breakthroughs [1–11] exemplified by the AlphaFold [1–3] and RoseTTAFold series [4, 5]. Despite these advances, accurate prediction of biomolecular complexes remains challenging.

To improve AlphaFold-like models of complexes, several specialized approaches have been proposed, such as denoising MSA profiles [12], integrating experimental constraints [13], or optimizing interaction interfaces through architectural advancements [14, 15]. Despite this progress, achieving consistent accuracy across diverse biomolecular systems remains a significant challenge. In parallel, improving the accuracy of a single prediction has proved difficult, and generating large ensembles of candidate conformations has therefore emerged as a simple yet powerful way to boost performance. The AlphaFold series improves precision through model ensembling, while AlphaFold3 [3] demonstrates that extensive sampling markedly enhances protein–antibody interface modeling. Recent studies [16–19] have generated thousands of candidates to expand conformational diversity. For instance, AFSample [16, 17] introduces stochasticity via dropout and random template masking to improve accuracy. Other strategies [20–23] diversify predictions through repeated MSA sub-sampling to probe alternative structural states.

Although generating large ensembles of candidate structures can improve predictive accuracy, it comes at substantial computational cost. For instance, performing a thousand samplings may require an entire GPU day. The main reason such massive sampling is needed is that current sampling methods are largely

unguided, leading to aimless exploration of conformational space. Specifically, predicting a biomolecular complex is essentially a search problem within a vast inter-chain conformational space, where only a few high-probability regions that correspond to potential binding modes are structurally plausible. The produced conformations often cluster around a single local region of the conformational landscape (Fig. 1a), producing redundant conformations that offer limited additional information. Consequently, sampling efficiency remains low. It is often unclear if further sampling is beneficial, as simpler targets gain little from extensive ensembling, rendering much computation redundant. These challenges highlight the need for planning-driven strategies that can first illustrate the most potentially informative regions, and then intelligently navigate the conformational space to focus computational effort where it matters most, thereby achieving higher accuracy at far lower computational cost.

Here, we introduce HelixFold-S1, a guided sampling strategy designed for biomolecular complexes that explores conformational space with exceptional efficiency and precision. In contrast to traditional stochastic search, HelixFold-S1 adopts a planning-based strategy to navigate the multi-modal landscape of complex structures. This approach strategically targets the most informative regions, which are conceptualized as high-probability modes within the conformational landscape, to facilitate the deliberate exploration of diverse and plausible binding modes (Fig. 1b). For each target complex, we first predict inter-chain inter-residue contact probabilities. Interpreting these distributions as a coarse-grained structural blueprint of the conformational space provides a reduced representation that highlights where meaningful interaction patterns are most likely to occur (Fig. 1c). This structural map then guides the sampling, with computational effort concentrated in regions associated with higher-probability, low-redundancy contacts that act as spatial constraints during structure generation. The resulting conformations are evaluated and ranked by their confidence scores, enabling both accurate and resource-efficient modeling of complex biomolecular assemblies (Fig. 1d).

We evaluated HelixFold-S1 across diverse biomolecular interfaces, including protein–protein, protein–antibody, protein–ligand, protein–RNA, and protein–DNA. Compared with conventional unguided sampling, HelixFold-S1 substantially improves prediction accuracy at equivalent sampling effort, with the largest gains observed for challenging protein–antibody complexes. Remarkably, it achieves comparable accuracy to traditional methods using an order of magnitude fewer sampling steps. The approach is generalizable to other folding models to enhance structural precision. Predicted inter-chain contact probabilities correlate strongly with structural difficulty and provide a practical guide for allocating sampling effort across targets. Guided sampling also improves exploration of the conformational space, producing more diverse and higher-quality ensembles.

Results

Guided Sampling Strategy and Architecture of HelixFold-S1

The inference pipeline of HF-S1 comprises two stages. In stage 1, HF-S1 predicts a coarse blueprint of the conformational space (Fig. 1c) to identify the most critical inter-chain interaction sites. An additional Contact Prediction Module (CPM) is employed to estimate inter-chain inter-token contact probabilities based on the pair representation, representing the likelihood that any two tokens are in spatial proximity, i.e., any atom pair lies within 5Å. This contact probability map provides a reduced, coarse-grained representation of the high-dimensional conformational space, serving as a structural blueprint to guide subsequent conformation sampling. In stage 2, conformations are generated guided by the predicted blueprint. Contacts are prioritized according to their predicted probabilities and sequentially selected from the blueprint. Each selected contact is introduced as an additional constraint via the Contact Conditioning Module (CCM), which processes these constraints and generates structures that attempt to satisfy them. This design aligns with conditional structure prediction [24–26], where geometric priors bias the model toward specific states. Our CCM uses the inter-chain contacts from Stage 1 as constraints to effectively steer subsequent structure prediction. By focusing on contacts with higher predicted probabilities, the model is more likely to produce multiple high-quality conformations that capture the most informative interactions. To reduce sampling redundancy, we employ a strategy called redundancy contact pruning (RCP). Once a conformation is generated using a particular contact, any contacts that are already satisfied within the resulting structure are excluded from further selection. All generated conformations are subsequently ranked according to model confidence scores, with the top-ranked prediction designated as the final output. Notably, RCP is an optional optimization; it enhances efficiency by sequentially pruning redundant contacts but can be bypassed for fully independent parallel generation. This allows HelixFold-S1 to balance high throughput via parallelism with refined, low-redundancy sampling, depending on available resources.

The architecture of HelixFold-S1 (HF-S1) builds upon that of HelixFold3 (HF3) [9], which reproduces the biomolecular structure prediction capabilities of AlphaFold3 (AF3) [3], and further extends its capacity to more intelligently explore conformational space through the addition of two contact-centric modules: the CPM and the CCM (Fig. 1e). A contact is considered to exist between two tokens if any pair of atoms from the corresponding tokens lies within 5Å. The CPM computes a contact probability matrix across all token pairs using the pair representations produced by the Pairformer, a specialized attention-based module designed to capture long-range pairwise token dependencies. The CCM, incorporated into the Input Embedder, enables the model to learn contact constraints and generate structures conditioned on these constraints.

During inference, the two modules are employed sequentially: in the first stage (Fig. 1c), the CPM estimates the overall contact landscape, and in the second stage (Fig. 1d), the CCM generates conformations guided by the predicted contacts. In contrast, during the training process, the model is optimized within a multi-task framework that alternates between two primary objectives: the contact prediction task (utilizing the CPM) and the contact-conditioned structure generation task (utilizing the CCM). This dual-task optimization allows the model to jointly learn inter-chain interactions and effectively leverage them during inference.

Improved Structural Accuracy across Complex Types

To systematically evaluate the performance of HF-S1, we constructed a test set comprising protein–antibody (n=221), protein–protein (n=198), protein–ligand (n=238), protein–RNA (n=177), and protein–DNA (n=254) complexes, collected between 2022.01.01 and 2024.12.31 from the RCSB PDB [27]. To minimize overlap with the training data, all test samples were selected to have low sequence identity to the training set. Sequences were clustered by similarity [28], and one representative per cluster was randomly chosen to ensure diversity and reduce redundancy. For protein–ligand complexes, any ligands that appeared in the training data were further excluded.

We compared AF3-like models , namely HF3 [9], Protenix [29], and Chai-1 [30], under different sampling strategies. For HF3, we examined two sampling-augmented variants: one following the AFSample strategy [16] (HF3 w/ AFSample) and another implementing our guided planning approach (HF-S1). To assess cross-model compatibility, we integrated HF-S1-predicted contacts into Protenix and Chai-1 (Protenix/Chai-1 w/ S1). This evaluation prioritizes sampling generalizability across architectures over absolute rankings. To manage the computational cost of benchmarking multiple external models, these variants were run in fully parallel mode without RCP; crucially, this trade-off does not compromise the fundamental assessment of our guided sampling strategy. Because HF-S1 may employ a distinct definition of contacts from those folding models, the baseline models may not reach their best performance under this evaluation. Moreover, Chai-1 currently supports contact constraints only for protein–protein complexes, and results are therefore reported for protein–antibody and protein–protein systems only. For all methods, we generated 25 predicted structures per target and ranked them using each method's confidence score.

Top-5 precision (Fig. 2a) reveals that HF-S1 consistently outperforms both HF3 and HF3 w/ AFSample. Improvements are most significant for protein–antibody complexes, exceeding 55% over HF3 and 33% over HF3 w/ AFSample. This is followed by protein–protein and protein–RNA/DNA interfaces, while protein–ligand binding sites show the smallest gains. This likely reflects the structural diversity of antigen-binding sites, which makes conventional sampling less effective at capturing native-like geometries, whereas HF-S1's guided sampling strategy increases the likelihood of generating correct interfacial contacts. For protein–protein and protein–RNA/DNA interfaces, HF-S1 provides moderate but clear improvements. By enforcing constraints on individual interfacial contacts, guided sampling helps capture correct local interactions, although single-contact constraints alone are not sufficient to fully determine the global geometry of these larger, more complex interfaces. Finally, protein–ligand binding sites are highly constrained, so while HF-S1 can still improve predictions, the relative gains are smaller compared with other complex types. Moreover, across different folding models, including HF3, Protenix, and Chai-1, incorporating the guided sampling strategy of HF-S1 leads to consistent performance gains, underscoring its broad effectiveness. In Fig. 2b, we further demonstrate that the performance advantage of HF-S1 over HF3 is consistently observed across different Top-K precisions of the sampled conformations. The improvements are pronounced for Top-1, Top-5, and Top-25, underscoring HF-S1's robustness across multiple ranking thresholds, although the Top-1 gains exhibit greater variability, potentially reflecting sensitivity to the selection of conformations based on the confidence score.

Extensive sampling is particularly effective for protein–antibody interfaces [29], and HF-S1 shows the largest gains on these targets. To systematically evaluate the benefit of extensive sampling on these interfaces,

we curated a dataset of 74 complexes released after 2024 and with residues less than 800 for efficiency, ensuring no overlap with Boltz-2's training data [31]. Sampling was conducted with top-performing methods evaluated 1,000 times per complex, while others were sampled 100 times per complex. Top-1 precision on protein–antibody interfaces (Fig. 2c) generally improves with increased sample size across all methods, particularly for those incorporating advanced strategies such as HF-S1 or AFsample. Importantly, HF-S1 achieves comparable precision to HF3 with AFsample using only 10 samples. This indicates that 10 HF-S1 samples are sufficient to reach the precision level that HF3 with AFsample requires 1,000 samples to achieve, representing just 1% of the computational cost. To further assess performance across different antibody types, Fig. 2d presents Top-1 precision for HF-S1 and HF3 with AFsample across canonical antibodies, nanobodies, and scFv based on the full 1,000-sample evaluation. HF-S1 demonstrates consistent improvements in Top-1 precision over HF3 w/ AFsample across all categories.

HF-S1 introduces two key components built on HF3: the Contact Prediction Module (CPM) and Contact Conditioning Module (CCM). To evaluate the CPM, we assessed the accuracy of its predicted contact probability matrices. For each target, token pairs corresponding to true contacts in experimental structures were treated as positives, and all others as negatives. We then computed the area under the precision–recall curve (AUPRC) between predicted and ground-truth contact maps, averaging across all targets. As a baseline, a posterior contact probability matrix for HF3 was derived by aggregating sampled conformations, with each entry defined as the inverse of the minimum inter-residue distance. Compared with HF3, HF-S1 consistently produced more accurate contact probabilities across diverse molecular types (Fig. 2e). Protein–antibody complexes showed the lowest AUPRC, reflecting the difficulty of this prediction scenario, whereas protein–ligand complexes achieved the highest AUPRC, indicating a relatively simpler task. This aligns with Fig. 2a: protein–antibody interfaces benefit most from guided sampling, while protein–ligand interfaces show smaller relative gains. To examine the CCM, we evaluated whether HF-S1–predicted conformations adhered to specified contact constraints. We defined the contact satisfaction rate as the fraction of predicted structures in which the given contacts were realized. Across most test cases, HF-S1 achieved satisfaction rates above 70% (Fig. 2f), demonstrating its ability to effectively incorporate contact priors into structure prediction.

Fig. 2g illustrates a representative case. Two inter-chain residue pairs (R43–N26 and V103–L28) were selected from HF-S1's predicted contact map and applied as spatial restraints during folding, yielding two distinct binding conformations that both satisfied the defined contact distances (R43–N26 $\approx$ 3.1 Å; V103–L28 $\approx$ 4.2 Å).

Contact Probability as an Indicator of Prediction Difficulty and Sampling Utility

The contact probability matrix predicted by HF-S1 serves a dual role: it not only informs the structural sampling strategy but also provides insight into the intrinsic difficulty of the structure prediction task, as well as the potential benefits of multiple sampling.

We first show that predicted contact probabilities can serve as a proxy for estimating target difficulty. Specifically, we defined the target-level contact probability as the maximum value in the predicted contact probability matrix for each target and analyzed its relationship with the Top-5 precision of HF-S1 using 25 samples (Fig. 3a). Targets were stratified into high, medium, and low groups, revealing a strong correlation across datasets: lower-probability targets consistently yielded poorer predictions, while higher-probability targets were predicted with high precision. This indicates that low contact probability generally corresponds to more difficult targets, which often require additional sampling to achieve accurate predictions. When grouped by complex type (Fig. 3b), a similar trend is observed: Protein–protein targets typically show higher target-level contact probabilities, while protein–antibody complexes cluster in the lower range, indicating greater structural uncertainty. Protein–ligand complexes consistently exhibit high contact probabilities, suggesting that HF-S1 can often localize ligand binding sites with high confidence. In contrast, protein–RNA and protein–DNA complexes display a broader distribution, reflecting greater variability in prediction difficulty across these categories.

We next examined whether target-level contact probabilities predict sampling utility. Targets were grouped based on their target-level predicted contact probability, and we analyzed the precision improvements of the best-performing multi-sample prediction relative to single-sample predictions (Fig. 3c). Targets with intermediate contact probabilities achieved the greatest improvements, while those with low or high probabilities saw more modest gains. Notably, although lower-probability targets did benefit from sampling, their improvements were generally smaller than those in the intermediate group. This trend is intuitive: for targets with high contact probabilities, accurate structures can often be recovered from a single sample,

leaving limited room for further enhancement. For lower-probability targets, the predicted contact maps are weak across the board, suggesting that a much larger number of samples may be needed to identify accurate structures. In contrast, intermediate cases offer partial yet informative contact signals, enabling the model to better explore the structural landscape and refine its predictions through sampling. As protein–ligand complexes consistently exhibit high target-level contact probabilities, they did not follow this trend.

We further examined performance across different contact probability groups by comparing HF-S1 (guided sampling) with HF3 (aimless sampling) (Fig. 3d). HF-S1 achieves the largest gains in the lower-probability group, while improvements are smaller in the higher-probability group. This trend suggests that targets with low contact probabilities are generally more challenging, and thus benefit more from guided exploration. In contrast, high-probability targets are easier to fold accurately, leaving less room for improvement.

Finally, we analyzed predicted contact probabilities across residue–residue pairs in the protein–protein dataset. Left panel of Fig. 3e: the averaged contact probability matrix across all residue pairs, with the pairs showing the highest probabilities indicated and those with the lowest probabilities highlighted. Right panel of Fig. 3e: compositional differences between the top 10 and bottom 10 pairs ranked by mean probability. Hydrophobic pairs such as Leu–Leu and Met-Met tend to show higher predicted probabilities, which could reflect their relatively frequent occurrence in tightly packed, energetically favorable environments. Conversely, polar or charged residues, such as Asp, Glu, and Lys, often exhibit lower probabilities, perhaps due to their general solvent exposure or context-specific interaction tendencies. These patterns indicate that contact prediction models may capture fundamental physicochemical preferences underlying residue interactions [32–35].

Guided Sampling Improves Exploration of Conformation Space

To test if guided sampling improves conformational exploration, we compared HF-S1 (guided planning) with HF3 (aimless sampling). We calculated the standard deviation (std) of precision scores per target as a proxy for structural diversity (Fig. 4a), where higher values indicate a broader exploration of the landscape. HF-S1 generally shows higher std values, with data concentrated in the upper-left region of the plot, reflecting more diverse ensembles. Targets were then grouped based on the difference in std between HF-S1 and HF3 (Fig. 4b): for protein–ligand complexes, a difference ≥ 1 indicates HF-S1 std higher, ≤ -1 indicates HF3 std higher, and intermediate values are comparable; for all other complex types, the threshold is 0.1. Across complex types, most targets fall into the HF-S1 std higher category, particularly protein–antibody, protein–RNA, and protein–DNA complexes, where the HF-S1 std higher category accounts for roughly twice as many targets as the HF3 std higher category. This enhanced diversity suggests that guided planning effectively directs sampling toward multiple plausible interaction geometries.

To examine how guided planning shapes the sampling trajectory, we tracked the average precision of all conformations accumulated up to each sampling step K (Fig. 4c). For HF3, precision remains nearly constant across steps, as the order of sampled conformations is effectively random. In contrast, HF-S1 exhibits a gradual decline in precision as sampling proceeds, reflecting its design to explore progressively less probable contact configurations. The higher precision observed in the early sampling stages arises from the use of higher-probability contacts as structural constraints, which guide the generation of more accurate conformations. Notably, HF-S1's initial conformations already outperform HF3 across most interface types, excluding protein–DNA complexes. This demonstrates that leveraging contact probabilities as structural constraints not only improves efficiency but also prioritizes structurally meaningful regions of the conformational landscape.

We compared the aimless sampling of HF3 with the guided strategy of HF-S1 using two representative examples (Fig. 4d): the Gel4–Nb4 nanobody complex (PDB ID: 8pe1) [36] and the Cas12m2–crRNA protein-RNA heterodimer (PDB ID: 8hhm) [37]. Analysis of predicted confidence and interface precision shows that HF-S1 consistently generates a broader spectrum of structures. These samples span a wide range of confidence and precision values, including a higher fraction of high-accuracy predictions. In contrast, HF3 samples cluster within narrower ranges, indicating more limited structural diversity and fewer high-quality candidates.

Discussion

Extensive sampling is standard for structural accuracy, yet unguided methods are computationally inefficient and lack insight into target difficulty. We propose that intelligent exploration of the conformational landscape enables efficient resource allocation. By directing sampling toward informative regions, our approach

enhances both efficiency and accuracy. Furthermore, predicted contact probabilities serve as a proxy for target difficulty, guiding users on the required sampling scale for convergence.

Beyond efficiency, this Search-then-Filter paradigm avoids local minima and consistently achieves superior Top-1 accuracy. For challenging targets, such as multi-epitope antibodies, cryptic pockets, or transitioning GPCRs, this prioritized ensemble ensures rare but biologically active conformations are not overlooked, providing value far beyond a single structure.

Despite these advances, assessing structural confidence remains challenging. Current metrics frequently misidentify the most accurate structures. High-confidence predictions may lack accuracy, while low-confidence sets often contain near-native conformations (Supplementary Figure 3). Refining these via sophisticated scoring or ensemble-based calibration could optimize candidate selection and prevent overlooking high-quality structures. Moreover, guided planning can exhibit diminishing returns. Greedy contact selection from highest to lowest probability may progressively sample less informative regions (Fig. 4c). Implementing early-stopping based on contact saturation or plateauing structural quality could sustain sampling efficiency. Finally, maintaining diversity is critical. Current sequential contact exclusion reduces redundancy but remains computationally serial. Although this strategy is effective (Supplementary Figure 4), more proactive methods are needed to preserve diversity more efficiently.

Intelligently navigating the conformational space remains a critical challenge. Sophisticated strategies, such as Monte Carlo Tree Search or reinforcement learning, could dynamically balance exploration and exploitation. Such methods promise to identify high-quality, diverse conformations more efficiently, enhancing the robustness of structure prediction across complex biomolecular targets.

Methods

Model Architecture

HF-S1 builds upon the HF3 architecture and is designed to support two complementary tasks: inter-chain contact prediction and contact-conditioned structure prediction. To this end, HF-S1 introduces two additional components: the Contact Prediction Module and the Contact Conditioning Module, which extend the base architecture to enable contact-level reasoning and constraint-based structure generation, respectively.

The contact prediction task aims to estimate the inter-chain contact distribution of a given protein complex. Various input features—including sequence, multiple sequence alignment (MSA), and template information—are first encoded and then processed by a Pairformer module to generate single and pair representations. These pairwise representations are subsequently passed to a dedicated Contact Prediction Module (CPM), which includes a Pairformer Stack and performs a binary classification task on each element of the pairwise representation to output contact probabilities for each inter-chain token pair. A contact is defined as the presence of any atom pair between two tokens within 5Å in 3D space. The resulting pairwise contact probability matrix captures the inter-chain contact distribution of the complex and serves as an informative intermediate representation. By operating in the simplified space of contacts—rather than directly in the complex, high-dimensional structural space—the model achieves greater computational efficiency and facilitates the contact-conditioned structure prediction task.

The contact-conditioned structure prediction task introduces a Contact Conditioning Module (CCM) to incorporate external contact constraints. These constraints are represented as a binary matrix $\{c_{ij}\}$, where $c_{ij} \in 0, 1$ indicates whether token pair (i, j) is in contact (1 if any atom pair is within 5Å, and 0 otherwise). This matrix is projected through a linear layer and then fused into the pairwise activations within the Input Embedder of the model. During training, for each complex, 0–10 inter-chain contacts are randomly sampled from the contact set extracted from its ground-truth structure and provided as input. During inference, contact constraints are sampled from the contact probability matrix produced by the contact prediction task; the specific sampling strategy is described in a later section. The model learns to utilize the provided contact information to enhance structure prediction accuracy.

Training Regime

Parameters of the newly introduced Contact Prediction Module and Contact Conditioning Module are randomly initialized, while the remaining parts of the model inherit weights from the pre-trained HF3. Fine-tuning is performed using the same training dataset as HelixFold3, which includes Protein Data Bank (PDB) [27] structures released before September 30, 2021, supplemented with self-distillation data to enhance generalization. The training follows a three-stage fine-tuning strategy: the first stage focuses on the

contact-conditioned structure prediction task to improve complex structure accuracy with inter-chain contact constraints; the second stage adds the contact prediction task, jointly optimizing the model for both tasks; the third stage follows the same setting of the second stage but extends to larger crop size.

In the first fine-tuning stage, the model is trained exclusively on the contact-conditioned structure prediction task. For each training sample, inter-chain contacts are extracted from experimentally determined complex structures to form a ground-truth contact set $\mathcal{C}$. This set contains all inter-chain token pairs where at least one atom from each token lies within 5Å in three-dimensional space. Contact conditioning is applied with 70% probability: 1–10 token pairs are uniformly sampled from $\mathcal{C}$ and provided as binary contact constraints. In the remaining 30% of samples, no contact constraints are used, which helps maintain the model's ability to predict structures without external guidance. While trained on individual experimental structures, the CPM learns underlying interaction probabilities by aggregating structural patterns from the vast diversity of homologous complexes and motifs in the training set. This allows the module to map a protein pair to a multi-modal probability landscape during inference, representing a global statistical distribution of plausible binding modes rather than a single deterministic state.

In the second and third fine-tuning stages, the model is optimized within a multi-task framework that simultaneously addresses both the contact-conditioned structure prediction task and the contact prediction task. The latter aims to estimate the probability that each inter-chain token pair is in atomic contact, serving as a basis for generating contact constraints during inference. During the specific optimization of the contact prediction task in these stages, the Contact Conditioning Module (CCM) is intentionally not activated; this is a deliberate design to ensure the CPM learns to infer intrinsic interaction patterns directly from raw features (MSA and sequences) without the "leakage" of ground-truth constraints. To supervise this task, a binary classification loss is applied over all inter-chain token pairs:

$$\mathcal{L}_{\text{contact}} = \frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} \text{cross\_entropy}(p_{ij}^{contact}, y_{ij}^{contact}).$$

Here, $\mathcal{P}$ denotes the set of all token pairs (i, j) such that token i and token j belong to different chains. $p_{ij}^{contact}$ is the predicted probability of contact between tokens i and j. $y_{ij}^{contact} = 1$ if $(i, j) \in \mathcal{C}$ (i.e., the token pair is in contact), and $y_{ij}^{contact} = 0$ otherwise. During training, half of the samples are used for the contact-conditioned structure prediction task, following the protocol established in the first stage, while the other half are dedicated to training the contact prediction task, which guides the model to estimate inter-chain contact probability distributions.

The loss function largely follows the original AF3/HF3 formulation, with an additional contact loss term introduced during fine-tuning:

$$\mathcal{L}_{\text{loss}} = \alpha_{\text{confidence}}\mathcal{L}_{\text{confidence}} + \alpha_{\text{diffusion}}\mathcal{L}_{\text{diffusion}} + \alpha_{\text{distogram}}\mathcal{L}_{\text{distogram}} + \alpha_{\text{contact}}\mathcal{L}_{\text{contact}},$$

with hyperparameters $\alpha_{\text{confidence}} = 0.01$, $\alpha_{\text{diffusion}} = 4$, and $\alpha_{\text{distogram}} = 0.3$. The contact loss coefficient α_{contact} is set to 1 during training samples used for the contact prediction task and 0 during samples used for the contact-conditioned structure prediction task. The definitions of all other loss terms remain consistent with those in AF3.

All stages use the Adam optimizer [38] with parameters $\beta_1 = 0.9$, $\beta_2 = 0.95$, and $\epsilon = 10^{-8}$, and a learning rate of 2×10^{-4}. The mini-batch size is fixed at 128 for all stages. The first fine-tuning stage consists of 10,000 training steps with a crop size of 384. The second fine-tuning stage extends to 20,000 steps, also with a crop size of 384. The third stage continues training for an additional 3,000 steps with an increased crop size of 640. We conducted model training on 128 Nvidia A100 GPUs, with a total duration of approximately 10 days.

Inference Regime

The inference process of HF-S1 (illustrated in Fig. 1a) consists of three stages: Contact Prediction, Contact Sampling, and Contact-Guided Structure Prediction and Ranking.

In the Contact Prediction stage, the contact prediction task is executed five times to reduce prediction variance, producing five contact probability matrices. These matrices are averaged element-wise to generate the final contact probability matrix, where each element $p_{ij}^{contact}$ represents the predicted contact probability between tokens i and j. Only inter-chain contact probabilities are retained, with intra-chain contacts set to zero. For protein–antibody complexes, contact sampling is performed exclusively between the antigen chain and each antibody chain (heavy and light), excluding contacts between heavy and light chains.

In the Contact Sampling stage, inter-chain contacts are selected sequentially in descending order according to their predicted contact probabilities. Each selected contact is used as a binary constraint in the subsequent structure prediction step to generate diverse candidate structures. To improve sampling efficiency and avoid redundancy, two strategies are adopted: redundant contact pruning and enriched sampling of previously identified contact sets. We denote the sets of contacts extracted from previously predicted structures as $C_1, C_2, \ldots$, where each C_k corresponds to the contacts obtained from the k-th predicted structure, following the same ground-truth extraction method described earlier. During sampling, redundant contact pruning excludes any candidate contact that overlaps with contacts already present in the union of all previously sampled sets $\bigcup_{i=1}^{k-1} C_i$. Here, overlapping means the candidate contact appears in any previously extracted contact set. This ensures that each newly sampled contact introduces novel constraints and helps maintain diversity among sampled structures. As sampling progresses, the predicted contact probabilities of remaining candidates naturally decrease. When these probabilities fall below a threshold (set as $0.2 \cdot \max_{i,j} p_{ij}^{contact}$), the benefit of exploring new low-confidence contacts diminishes. At this point, instead of sampling new contacts, the algorithm enriches sampling by iterating through the existing contact sets $C_1, C_2, \ldots$ in order. Contacts are drawn from these sets cyclically to further exploit high-confidence information until the total sampling budget S is reached.

In the Contact-Guided Structure Prediction and Ranking stage, each sampled contact is treated as a binary constraint and passed into the contact-conditioned structure prediction task to generate a candidate structure. A confidence score, named ranking_confidence, is computed for each structure, and the final prediction is selected as the one with the highest confidence among all candidates. Drawing inspiration from the AF3, we define the confidence score as a weighted average of the pTM and ipTM scores, with an additional penalty term for structural clashes. The score is computed as follows:

$$\text{ranking\_confidence} = 0.2 \cdot \text{pTM} + 0.8 \cdot \text{ipTM} - 1.0 \cdot \text{has\_clash},$$

where pTM represents the predicted TM-Score for the full complex, indicating the confidence for overall structural accuracy. ipTM represents the interface predicted TM-Score for the full complex, focusing on the accuracy of interfacial interactions. has_clash is a binary term indicating the presence of obvious clashes between polymer chains in the predicted structure. Detailed definitions of pTM, ipTM, and has_clash can be found in the AF3 paper [3].

We adopt consistent inference settings across structure prediction tasks, including our method and the baselines Boltz-2[31], Protenix[29], and Chai-1[30]. Each prediction is refined using 10 recycling iterations and 200 diffusion steps, where the diffusion module is run once to generate a single structure per input. In the corresponding figures, lines represent these average outcomes, while shaded areas indicate the variability between the two runs. Notably, the inference configuration for HF3 w/ AFsample adopts a more sophisticated multi-setting approach, according to the AlphaFold settings used in AFsample [16]. The complete inference specifications for HF3 with AFsample integrate three distinct hyperparameter settings as detailed in Table 1. Protenix w/ S1, Chai-1 w/ S1, and Boltz-2 w/ S1 follow the same workflow as HF-S1, differing only in the backbone structure prediction module employed during the final stage. Contact probability outputs from HF-S1 are ranked from highest to lowest, and the corresponding contact pairs are incorporated as external geometric constraints to guide structure generation. It should be noted that Chai-1 w/ S1 is restricted to protein–protein/antibody datasets, as it only accepts residue-residue contact constraints as input. All methods construct MSAs using their respective built-in sequence search tools.

Evaluation Data

Evaluation sets for protein–protein, protein–ligand, protein–RNA, and protein–DNA interfaces were constructed from all PDB entries released between May 1, 2022 and December 31, 2024, with each structure expanded to Biological Assembly 1. Interfaces were defined as pairs of entities with a minimum heavy-atom distance below 5 Å. Protein–antibody complexes were sourced from SAbDab [39] within the same date range, using symmetric units instead of Biological Assembly 1.

For targets collected from the PDB, complexes with resolution worse than 4.5 Å or exceeding 1400 tokens under our tokenization scheme were removed. Polymer–polymer interfaces were excluded if both polymers shared more than 40% sequence identity with two chains from the same PDB entry in the training set. For protein–ligand interfaces, the following criteria were applied: (1) only ligands with CCD codes absent from the training set were retained; (2) covalently bound ligands, including those involved in glycosylation, were excluded; (3) ligands containing five or fewer atoms or occurring in ten or more PDB entries were removed; (4) only ligands with molecular weights between 100 and 900 Da were retained; (5) ligands were required to exhibit a *ranking_model_fit* score of at least 0.5, as reported in the RCSB structure validation dataset,

indicating above-median model quality for X-ray crystallographic structures [40]; and (6) binding pockets were required to include between 5 and 100 protein residues within 5 Å of the ligand.

We clustered the remaining targets by grouping proteins with nine or more residues at 40% sequence identity, while nucleic acids and proteins with nine or fewer residues were clustered at 100%, using MMseqs2 with a minimum coverage of 80% and default clustering mode. Each interface was assigned a binary, order-independent cluster ID based on entity pairs—(polymer1_cluster, polymer2_cluster) for polymer–polymer interfaces and (polymer_cluster, ligand_CCD-code) for protein–ligand interfaces. Evaluation was performed on one representative entry per cluster.

Protein–antibody complexes were sourced from SAbDab [39], including only those with resolution better than 9 Å, and containing antigen chains. The antigen sequences were grouped into clusters based on a 40% sequence identity threshold. Subsequently, we retained only those clusters that contained no cases released prior to September 30, 2021. Further filtering was applied to select cases released within the period from May 1, 2022, to December 31, 2024. Ultimately, one case was chosen from each of the remaining clusters to constitute the evaluation set for protein-antibody analysis. For efficiency considerations in the extensive sampling test(as shown in Fig. 2c), we exclusively selected samples released between January 1, 2024, and December 31, 2024, and filtered out those with over 800 residues, ultimately retaining 74 samples.

Evaluation Metrics

To evaluate structure prediction performance across different interaction types, we adopt distinct metrics tailored to the characteristics of each molecular interface.

Protein–protein complexes, including protein–antibody interactions, are evaluated using DockQ [41], which integrates interface RMSD, FNAT, and FNAS to provide a reliable summary of interface quality. For protein–antibody complexes specifically, all antibody chains are treated collectively as the "ligand", and DockQ is computed over the interface between the antibody and the rest of the complex using the DockQ v1 implementation.

Nucleic acid–protein interfaces, including both protein–RNA and protein–DNA complexes, are assessed using interface LDDT (iLDDT) [42], computed over atom pairs across different chains within a 30Å inclusion radius to accommodate the larger and more diffuse interaction footprints characteristic of nucleic acids.

Protein–ligand complexes are evaluated using pocket-aligned RMSD, which measures ligand pose accuracy after aligning the predicted structure to the binding pocket of the ground truth. The pocket is defined as all heavy atoms within 10Å of any heavy atom of the ligand in the ground truth structure, restricted to the primary protein chain—identified as the chain containing the most atoms within this radius. The C^{α} atoms of this pocket are used to perform a least-squares alignment between predicted and reference structures. After alignment, a symmetry-corrected ligand RMSD is computed over all heavy atoms of the ligand using RDKit's Chem.rdMolAlign.CalcRMS [43], which aligns the ligands while accounting for molecular symmetry before computing the final deviation.

The precision of the Top-K ranked structures is defined as the highest accuracy achieved among the top K structures ordered by their confidence score.

Data Availability

To train HelixFold-S1, PDB can be downloaded at https://www.rcsb.org/docs/programmatic-access/file-download-services and AlphaFold Protein Structure Database as the distillation dataset can be downloaded at https://ftp.ebi.ac.uk/pub/databases/alphafold/v2/. The test set are filtered and clustered from PDB with conditions detailed in Methods. The protein-antibody complexes for test can be downloaded at https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab/. Detailed data processing procedures, including filtering, clustering, and dataset construction, are described in the Methods section.

Code Availability

The source code, trained weights, and inference scripts for HelixFold-S1 are publicly available at https://github.com/PaddlePaddle/PaddleHelix/tree/dev/apps/protein_folding/HelixFold-S1. To ensure long-term availability, the version of the code at the time of acceptance has been archived on Zenodo (DOI: https://doi.org/10.5281/zenodo.8202943) [44].

Acknowledgments

This work was supported by the National Science and Technology Major Project (2023ZD0120803).

Author Contributions

X. Fang and X. Zhang led the research. L. Liu and X. Fang contributed technical ideas. L. Liu, Y. Liu, and X. Ye developed the proposed method. S. Zhang, Y. Li, K. Zhu, Y. Xue, and J. Zhou developed analytics. L. Liu, X. Fang, Y. Liu, and X. Ye wrote the paper.

Competing Interests

The authors declare no competing interests.

Settings	Templates	Dropout	Recycles	Ratio %
setting-1	Yes	Yes	3	30
setting-2	No	Yes	3	30
setting-3	No	Yes	9	40

Table 1 Inference configurations of HF3 w/ AFsample. The term *Templates* indicates whether structural templates were employed. *Dropout* denotes whether the dropout mechanism was activated. *Recycles* signifies the number of recycling operations utilized, with a default value of 3. *Ratio* represents the proportion that this particular setting occupies within the entire sampling process.

Fig. 1— **Overall framework of HelixFold-S1. a, b,** Conceptual comparison of sampling paradigms: **a,** Aimless exploration in conventional conformation sampling, often concentrated within limited regions of the conformational space; **b,** Guided exploration in planning-based sampling, illustrating the strategic advantage of directing sampling toward diverse conformational states. **c, d,** Two-stage inference of HelixFold-S1: **c,** stage 1 predicts a blueprint of the conformational space; **d,** stage 2 generates conformations based on this blueprint. **e,** Network architecture of HF-S1, which incorporates a contact conditioning module and a contact prediction module into the HelixFold3's model architecture.

Fig. 2— **Structural performance and module evaluation of HelixFold-S1 across multiple complex types. a,** Top-5 structural precision among 25 sampled conformations for various folding models. Benchmark complexes were collected from the RCSB PDB between January 1, 2022, and December 31, 2024, including protein–antibody ($n = 221$), protein–protein ($n = 198$), protein–ligand ($n = 238$), protein–RNA ($n = 177$), and protein–DNA ($n = 254$) complexes. Statistical significance was determined using a two-sided paired t-test. On the protein-protein dataset, the difference between HF-S1 and HF3 was significant ($p \approx 0.0029$); for all other datasets, the improvements were highly significant ($p < 10^{-6}$). **b,** Relative improvements of HelixFold-S1 over HelixFold3 across different complex types in Top-1, Top-5, and Top-25 precision. **c,** Comparison of HF-S1 with various baseline models on protein–antibody complexes released in 2024 (n=74), showing that extensive sampling progressively improves structural accuracy. Some methods were sampled 1000 times, while others used 100 samples. **d,** Performance comparison between HF3 w/ AFSample and HF-S1 across different antibody types on the same 2024 protein–antibody complexes (n=74), using 1000 samples. **e,** Performance of the contact prediction module in HF-S1, evaluated using the area under the precision–recall curve (AUPRC). **f,** Effectiveness of the contact conditioning module, measured as the fraction of predicted structures satisfying contact constraints. In the box plots of **e** and **f**, the center line indicates the median; the box limits represent the 25th and 75th percentiles; and the whiskers extend to the maximum and minimum values, excluding outliers. **g,** Example structures predicted by HelixFold-S1 using two contact constraints (PDB ID: 8ozb).

Fig. 3— **Contact probability as an indicator of prediction difficulty and sampling utility. a,** Correlation between target-level contact probability (maximum value in the predicted matrix) and structural precision across interface types, evaluated using the Top-5 of 25 sampled conformations. In the box plots, the center line indicates the median; the box limits represent the 25th and 75th percentiles; and the whiskers extend to the minimum and maximum values (excluding outliers). **b,** Distribution of predicted contact probabilities across interface types, constructed from all benchmark targets in each category. **c,** Structural improvement from single to extensive sampling by HF-S1 across contact probability groups, measured using the best-performing conformations, showing larger gains for lower-probability interfaces. **d,** Relative Top-5 precision (of 25 sampled conformations) of HF-S1 compared with HF3 across contact probability levels, indicating stronger benefits on lower-probability targets. The sample numbers in this analysis match those in **a** and **c**. **e,** Predicted contact probabilities reflect residue physicochemical preferences.

Fig 4— **Guided sampling improves exploration of conformational space. a,** Target-level standard deviation of structural precision among sampled conformations for HF3 and HF-S1. Each point represents a target; higher values indicate greater conformational diversity. **b,** Fraction of targets with higher standard deviation for HF-S1, higher for HF3, or comparable, summarized across different interface types.**c,** Structural precision as a function of the number of sampled conformations. **d,** Distribution of ranking confidence and structural precision of sampled structures for two representative complexes (PDB: 8PE1 and 8HHM). HF-S1 shows a broader distribution of structural precision, whereas HF3 predictions are more concentrated. Right panels show the corresponding predicted structures by HF-S1 and HF3 compared to experimental structures.

References

[1] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., *et al.*: Highly accurate protein structure prediction with alphafold. nature **596**(7873), 583–589 (2021)

[2] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al.: Protein complex prediction with alphafold-multimer. biorxiv, 2021–10 (2021)

[3] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al.: Accurate structure prediction of biomolecular interactions with alphafold 3. Nature, 1–3 (2024)

[4] Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., *et al.*: Accurate prediction of protein structures and interactions using a three-track neural network. Science **373**(6557), 871–876 (2021)

[5] Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G.R., Morey-Burrows, F.S., Anishchenko, I., Humphreys, I.R., *et al.*: Generalized biomolecular modeling and design with rosettafold all-atom. Science **384**(6693), 2528 (2024)

[6] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., *et al.*: Evolutionary-scale prediction of atomic-level protein structure with a language model. Science **379**(6637), 1123–1130 (2023)

[7] Fang, X., Wang, F., Liu, L., He, J., Lin, D., Xiang, Y., Zhu, K., Zhang, X., Wu, H., Li, H., *et al.*: A method for multiple-sequence-alignment-free protein structure prediction using a protein language model. Nature Machine Intelligence **5**(10), 1087–1096 (2023)

[8] Fang, X., Gao, J., Hu, J., Liu, L., Xue, Y., Zhang, X., Zhu, K.: Helixfold-multimer: Elevating protein complex structure prediction to new heights. arXiv preprint arXiv:2404.10260 (2024)

[9] Liu, L., Zhang, S., Xue, Y., Ye, X., Zhu, K., Li, Y., Liu, Y., Gao, J., Zhao, W., Yu, H., et al.: Technical report of helixfold3 for biomolecular structure prediction. arXiv preprint arXiv:2408.16975 (2024)

[10] Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M.: Colabfold: making protein folding accessible to all. Nature methods **19**(6), 679–682 (2022)

[11] Hayes, T., Rao, R., Akin, H., Sofroniew, N.J., Oktay, D., Lin, Z., Verkuil, R., Tran, V.Q., Deaton, J., Wiggert, M., et al.: Simulating 500 million years of evolution with a language model. Science, 0018 (2025)

[12] Bryant, P., Noé, F.: Improved protein complex prediction with alphafold-multimer by denoising the msa profile. PLOS Computational Biology **20**(7), 1012253 (2024)

[13] Mirabello, C., Wallner, B., Nystedt, B., Azinas, S., Carroni, M.: Unmasking alphafold to integrate experiments and predictions in multimeric complexes. Nature Communications **15**(1), 8724 (2024)

[14] Bryant, P., Pozzati, G., Elofsson, A.: Improved prediction of protein-protein interactions using alphafold2. Nature communications **13**(1), 1265 (2022)

[15] Akiyama, Y., Zhang, Z., Mirdita, M., Steinegger, M., Ovchinnikov, S.: Scaling down protein language modeling with msa pairformer. bioRxiv, 2025–08 (2025)

[16] Wallner, B.: Afsample: improving multimer prediction with alphafold using massive sampling. Bioinformatics **39**(9), 573 (2023)

[17] Kalakoti, Y., Wallner, B.: Afsample2 predicts multiple conformations and ensembles with alphafold2. Communications Biology **8**(1), 373 (2025)

[18] Stein, R.A., Mchaourab, H.S.: Speach_af: Sampling protein ensembles and conformational heterogeneity

with alphafold2. PLOS Computational Biology **18**(8), 1010483 (2022)

[19] Yin, R., Pierce, B.G.: Evaluation of alphafold antibody–antigen modeling with implications for improving predictive accuracy. Protein Science **33**(1), 4865 (2024)

[20] Xing, E., Zhang, J., Wang, S., Cheng, X.: Leveraging sequence purification for accurate prediction of multiple conformational states with alphafold2. Research Square, 3 (2025)

[21] Silva, G., Cui, J.Y., Dalgarno, D.C., Lisi, G.P., Rubenstein, B.M.: High-throughput prediction of protein conformational distributions with subsampled alphafold2. nature communications **15**(1), 2464 (2024)

[22] Wayment-Steele, H.K., Ojoawo, A., Otten, R., Apitz, J.M., Pitsawong, W., Hömberger, M., Ovchinnikov, S., Colwell, L., Kern, D.: Predicting multiple conformations via sequence clustering and alphafold2. Nature **625**(7996), 832–839 (2024)

[23] Bryant, P., Noé, F.: Structure prediction of alternative protein conformations. Nature Communications **15**(1), 7328 (2024)

[24] Stahl, K., Graziadei, A., Dau, T., Brock, O., Rappsilber, J.: Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. Nature Biotechnology **41**(12), 1810–1819 (2023)

[25] Stahl, K., Warneke, R., Demann, L., Bremenkamp, R., Hormes, B., Brock, O., Stülke, J., Rappsilber, J.: Modelling protein complexes with crosslinking mass spectrometry and deep learning. Nature communications **15**(1), 7866 (2024)

[26] Heo, L., Feig, M.: Multi-state modeling of g-protein coupled receptors at experimental accuracy. Proteins: Structure, Function, and Bioinformatics **90**(11), 1873–1885 (2022)

[27] Burley, S.K., Berman, H.M., Kleywegt, G.J., Markley, J.L., Nakamura, H., Velankar, S.: Protein data bank (pdb): the single global macromolecular structure archive. Protein crystallography: methods and protocols, 627–641 (2017)

[28] Steinegger, M., Söding, J.: Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nature biotechnology **35**(11), 1026–1028 (2017)

[29] Team, B.A.A., Chen, X., Zhang, Y., Lu, C., Ma, W., Guan, J., Gong, C., Yang, J., Zhang, H., Zhang, K., et al.: Protenix-advancing structure prediction through a comprehensive alphafold3 reproduction. BioRxiv, 2025–01 (2025)

[30] team, C.D., Boitreaud, J., Dent, J., McPartlon, M., Meier, J., Reis, V., Rogozhonikov, A., Wu, K.: Chai-1: Decoding the molecular interactions of life. BioRxiv, 2024–10 (2024)

[31] Passaro, S., Corso, G., Wohlwend, J., Reveiz, M., Thaler, S., Ram Somnath, V., Getz, N., Portnoi, T., Roy, J., Stark, H., et al.: Boltz-2: Towards accurate and efficient binding affinity prediction. BioRxiv, 2025–06 (2025)

[32] Yan, C., Wu, F., Jernigan, R.L., Dobbs, D., Honavar, V.: Characterization of protein–protein interfaces. The Protein Journal **27**(1), 59–70 (2008)

[33] Kastritis, P.L., Bonvin, A.M.J.J.: On the binding affinity of macromolecular interactions: daring to ask why proteins interact. Journal of The Royal Society Interface **10**(79), 20120835 (2013)

[34] Zhou, H.-X., Pang, X.: Electrostatic interactions in protein structure, folding, binding, and condensation. Chemical Reviews **118**(4), 1691–1741 (2018). PMID: 29319301

[35] Seychell, B.C., Beck, T.: Molecular basis for protein–protein interactions. Beilstein Journal of Organic Chemistry **17**, 1–10 (2021)

[36] Redrado-Hernández, S., Macías-León, J., Castro-López, J., Belén Sanz, A., Dolader, E., Arias, M., González-Ramírez, A.M., Sánchez-Navarro, D., Petryk, Y., Farkaš, V., *et al.*: Broad protection against

invasive fungal disease from a nanobody targeting the active site of fungal β-1, 3-glucanosyltransferases. Angewandte Chemie **136**(34), 202405823 (2024)

[37] Omura, S.N., Nakagawa, R., Südfeld, C., Villegas Warren, R., Wu, W.Y., Hirano, H., Laffeber, C., Kusakizako, T., Kise, Y., Lebbink, J.H., *et al.*: Mechanistic and evolutionary insights into a type vm crispr–cas effector enzyme. Nature structural & molecular biology **30**(8), 1172–1182 (2023)

[38] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

[39] Schneider, C., Raybould, M.I.J., Deane, C.M.: SAbDab in the age of biotherapeutics: updates including SAbDab-nano, the nanobody structure tracker. Nucleic Acids Research **50**, 1368–1372

[40] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The Protein Data Bank. Nucleic Acids Research **28**(1), 235–242 (2000)

[41] Basu, S., Wallner, B.: Dockq: A quality measure for protein-protein docking models. PLOS ONE **11**(8), 1–9 (2016)

[42] Mariani, V., Biasini, M., Barbato, A., Schwede, T.: lddt: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics **29**(21), 2722–2728 (2013)

[43] Landrum, G.: RDKit: Open-source Cheminformatics. http://www.rdkit.org

[44] Fang, X. et al.: PaddleHelix: HelixFold-S1 Source Code and Inference Scripts. Zenodo. https://doi.org/10.5281/zenodo.8202943

Editor Summary:

Liu and colleagues introduce HelixFold-S1, a guided sampling strategy for biomolecular complex structure predicition that targets high-probability interaction regions. The method achieves higher accuracy than traditional unguided methods while reducing computational costs.

Peer Review Information:

*Nature Machine Intelligenc*e thanks Sonya M. Hanson, and the other, anonymous, reviewers for their contribution to the peer review of this work.

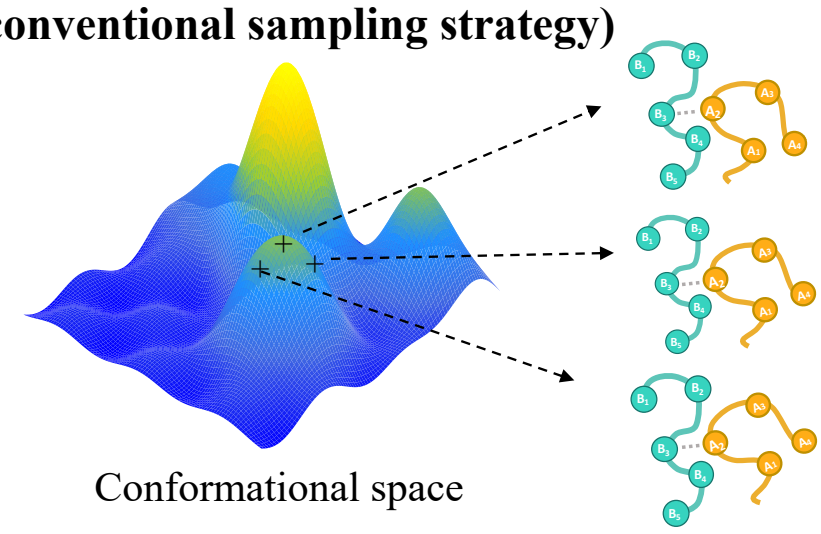
1. Supplementary Information:

A. PDF Files

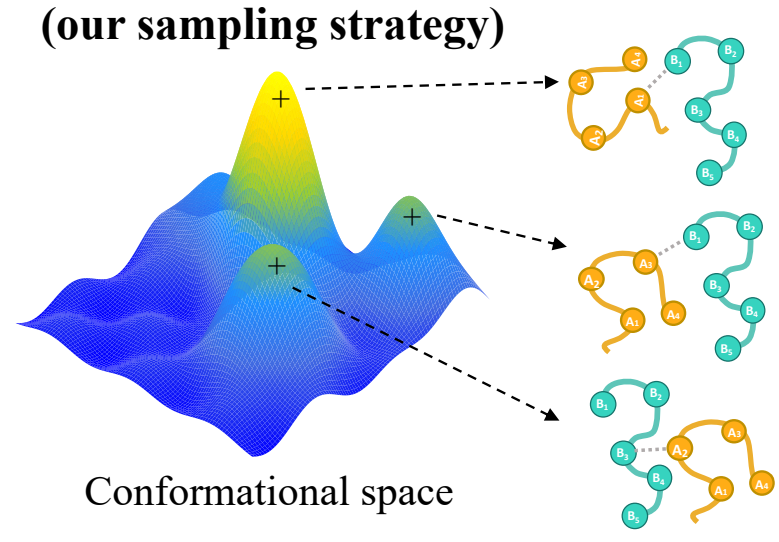
Item	**Present?**	**Filename** **Whole original file name including extension. i.e.: Smith_SI.pdf. The extension must be .pdf**	**A brief, numerical description of file contents.** **i.e.: *Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.***
Supplementary Information	Yes	supplementary_information.pdf	Supplementary Fig.1-12
Reporting Summary	Yes	reporting_summary.pdf	
Peer Review Information	Yes	*OFFICE USE ONLY*	

Conceptual Comparison

a **Aimless sampling**
(conventional sampling strategy)



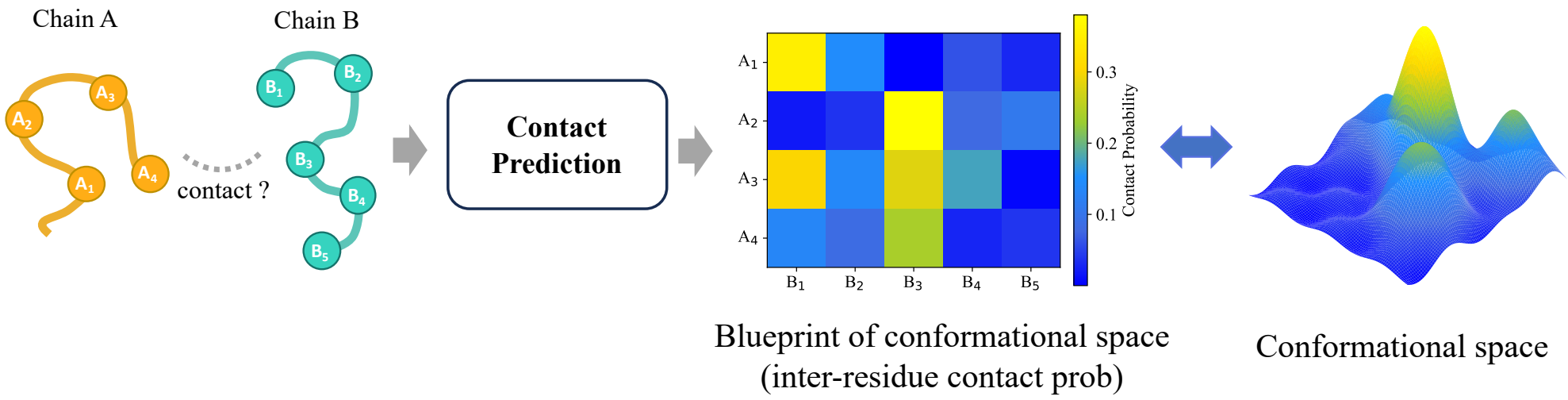
b **Guided planning**
(our sampling strategy)



Inference Pipeline

c

Stage 1: Predict blueprint of conformational space

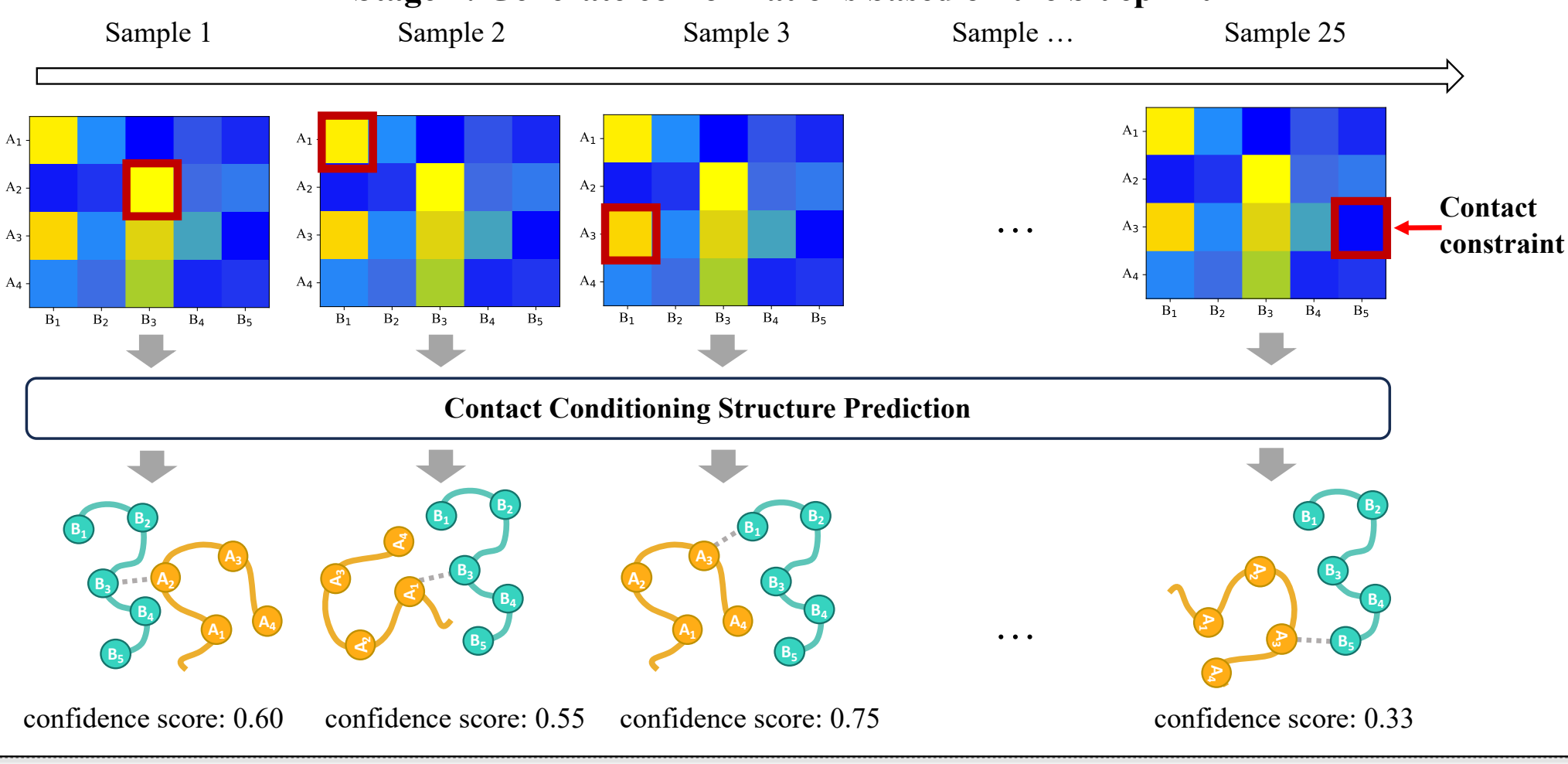


Blueprint of conformational space
(inter-residue contact prob)

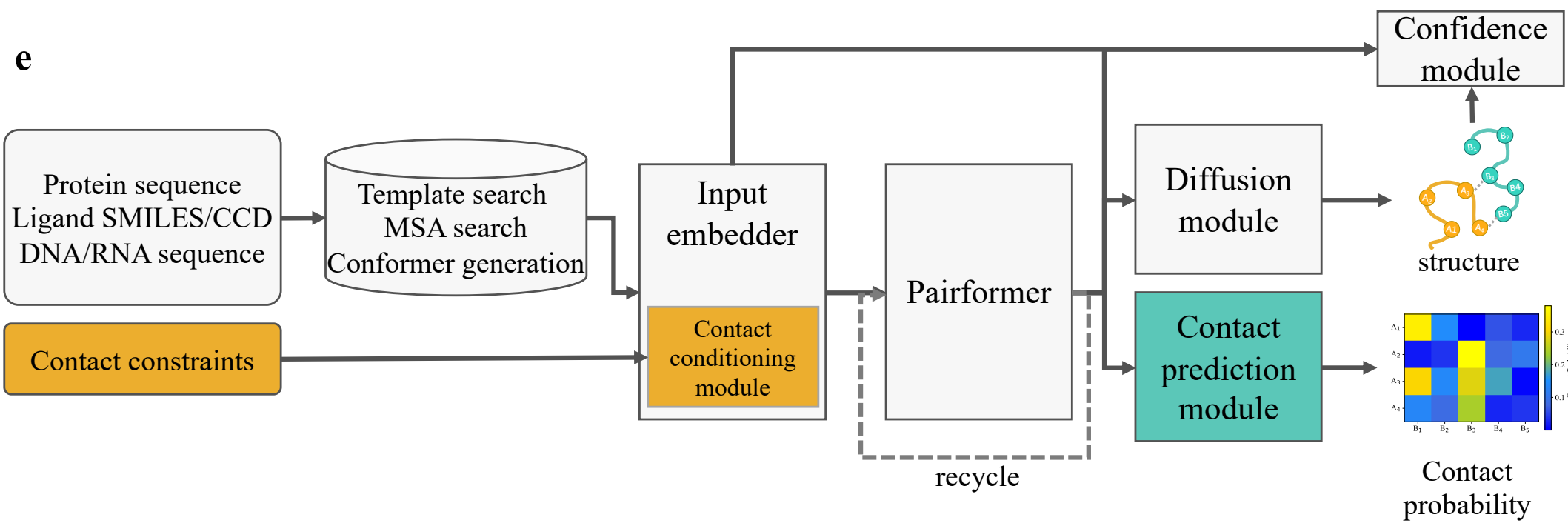
Conformational space

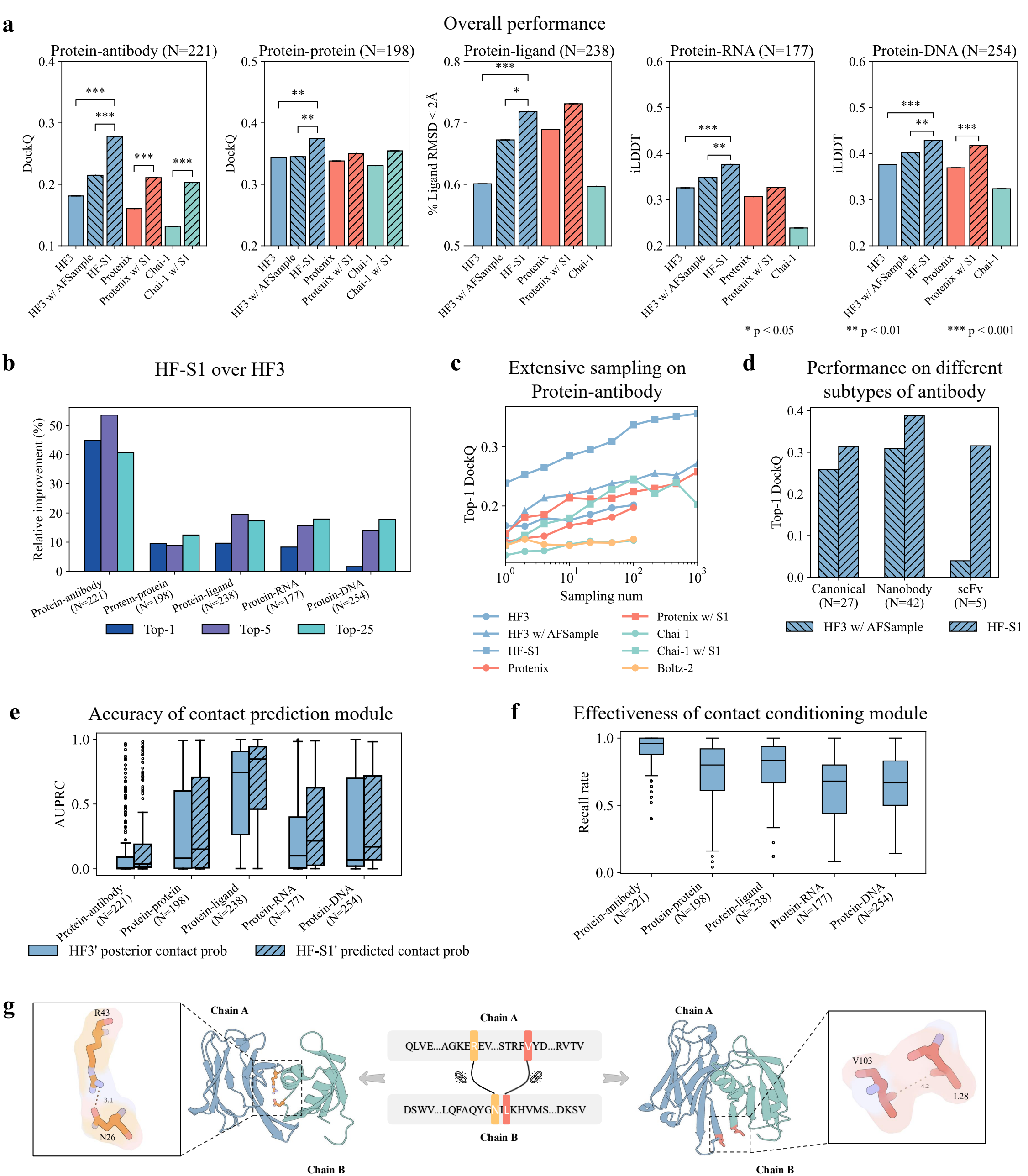
d

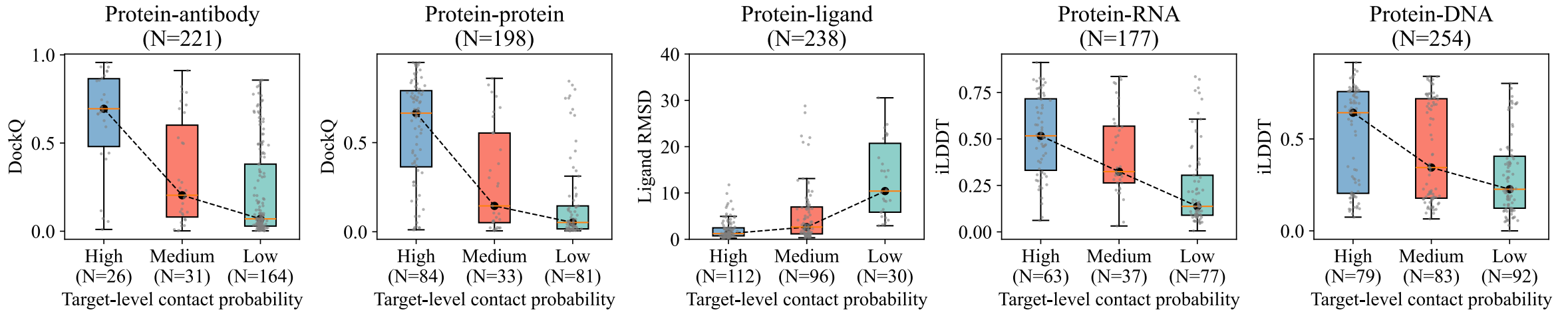
Stage 2: Generate conformations based on the blueprint

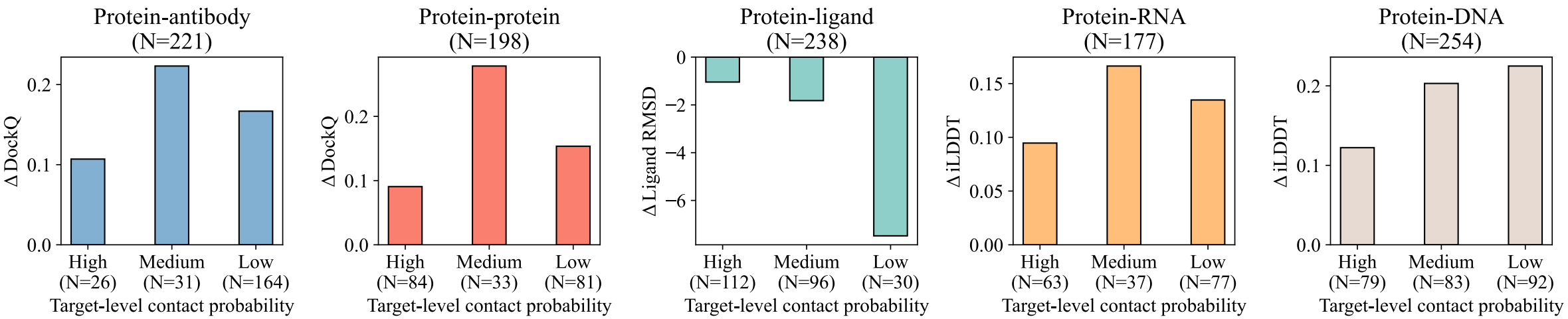


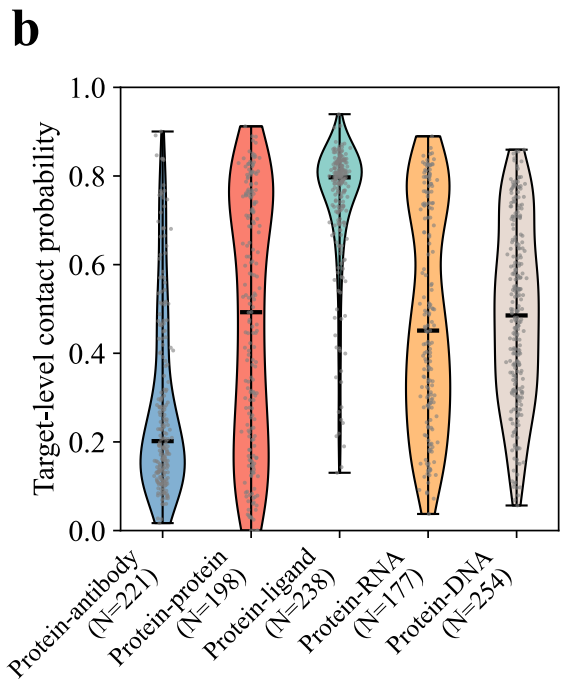
e

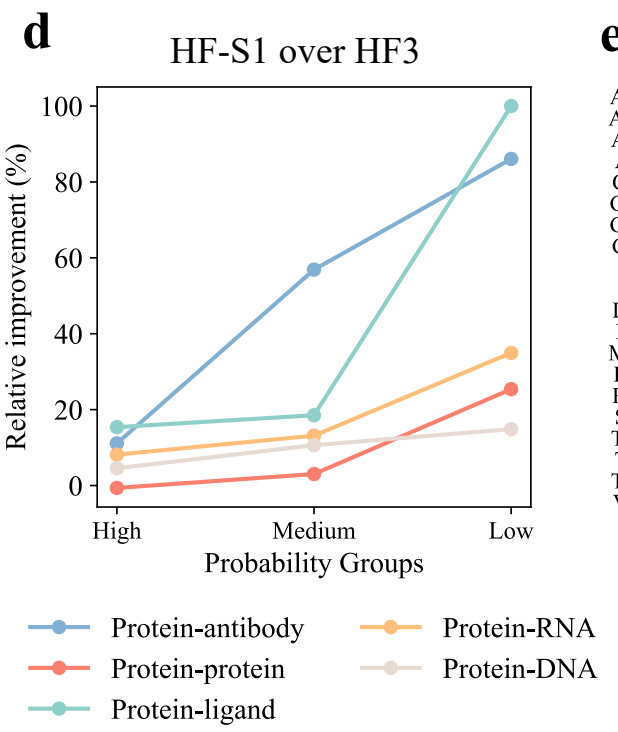


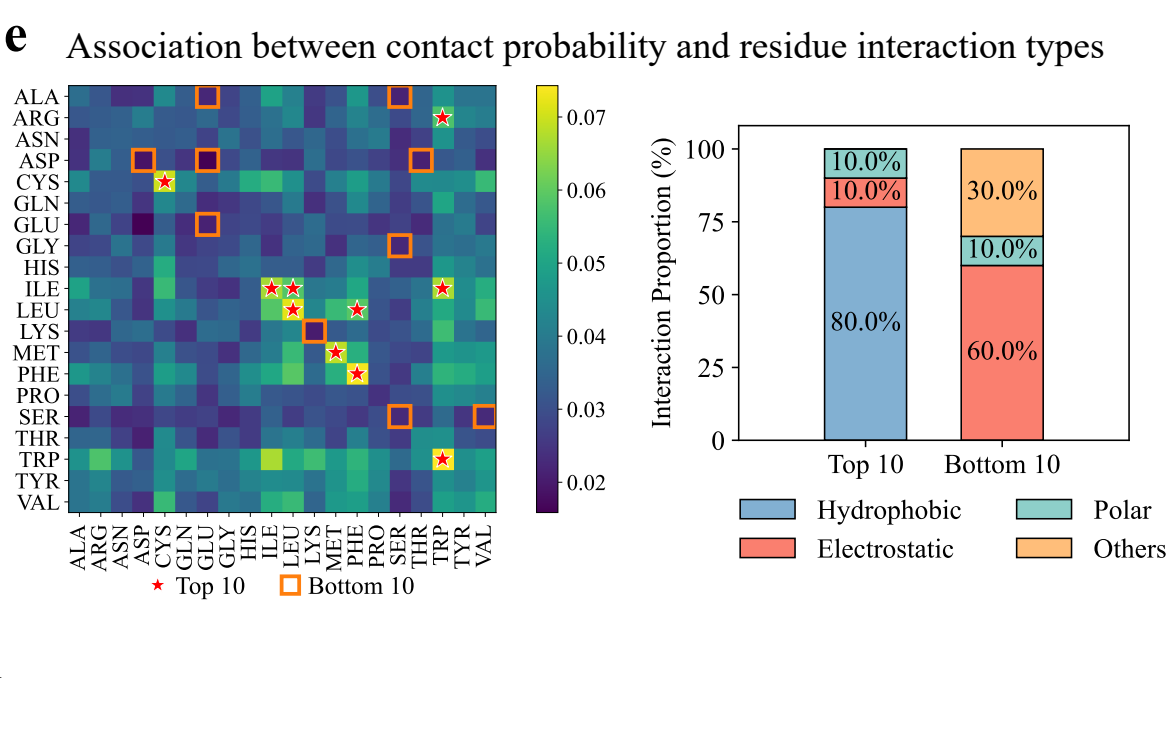

a
Overall performance
Protein-antibody (N=221)
Protein-protein (N=198)
Protein-ligand (N=238)
Protein-RNA (N=177)
Protein-DNA (N=254)
DockQ
% Ligand RMSD < 2Å
iLDDT
HF3
HF3 w/ AFSample
HF-S1
Protenix
Protenix w/ S1
Chai-1
Chai-1 w/ S1
* p < 0.05
** p < 0.01
*** p < 0.001
b
HF-S1 over HF3
Relative improvement (%)
Protein-antibody (N=221)
Protein-protein (N=198)
Protein-ligand (N=238)
Protein-RNA (N=177)
Protein-DNA (N=254)
Top-1
Top-5
Top-25
c
Extensive sampling on Protein-antibody
Top-1 DockQ
Sampling num
HF3
HF3 w/ AFSample
HF-S1
Protenix
Protenix w/ S1
Chai-1
Chai-1 w/ S1
Boltz-2
d
Performance on different subtypes of antibody
Top-1 DockQ
Canonical (N=27)
Nanobody (N=42)
scFv (N=5)
HF3 w/ AFSample
HF-S1
e
Accuracy of contact prediction module
AUPRC
HF3' posterior contact prob
HF-S1' predicted contact prob
f
Effectiveness of contact conditioning module
Recall rate
g
R43
N26
3.1
Chain A
Chain B
QLVE...AGKEREV...STRFVYD...RVTV
DSWV...LQFAQYGNILKHVMS...DKSV
V103
L28
4.2

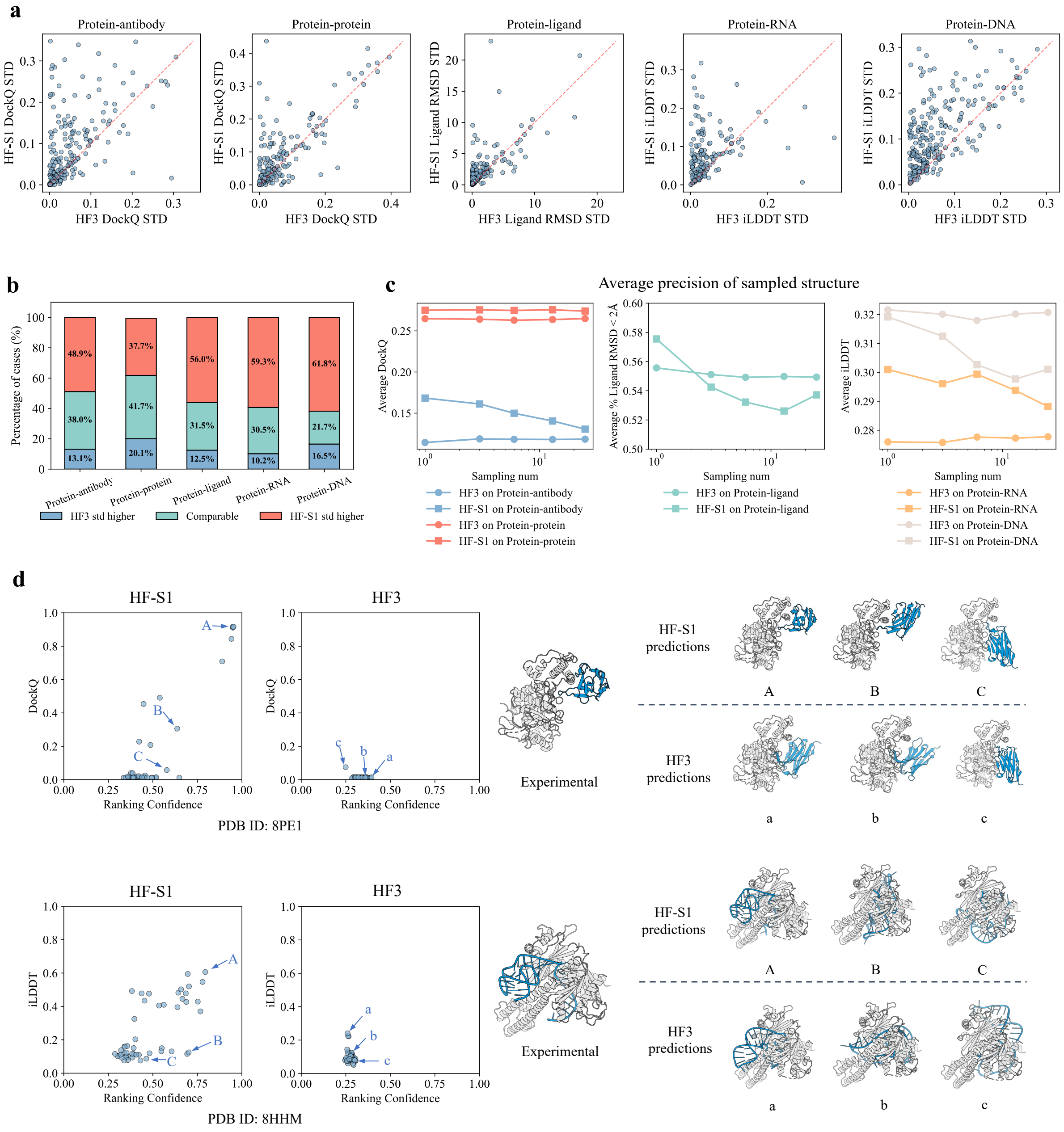

a
Correlation between target-level contact probability and structural precision
Protein-antibody (N=221)
Protein-protein (N=198)
Protein-ligand (N=238)
Protein-RNA (N=177)
Protein-DNA (N=254)
DockQ
Ligand RMSD
iLDDT
High (N=26) Medium (N=31) Low (N=164)
High (N=84) Medium (N=33) Low (N=81)
High (N=112) Medium (N=96) Low (N=30)
High (N=63) Medium (N=37) Low (N=77)
High (N=79) Medium (N=83) Low (N=92)
Target-level contact probability


c
Improvement of multiple sampling compared with single sampling
Protein-antibody (N=221)
Protein-protein (N=198)
Protein-ligand (N=238)
Protein-RNA (N=177)
Protein-DNA (N=254)
ΔDockQ
ΔLigand RMSD
ΔiLDDT
High (N=26) Medium (N=31) Low (N=164)
High (N=84) Medium (N=33) Low (N=81)
High (N=112) Medium (N=96) Low (N=30)
High (N=63) Medium (N=37) Low (N=77)
High (N=79) Medium (N=83) Low (N=92)
Target-level contact probability


b
Target-level contact probability
Protein-antibody (N=221)
Protein-protein (N=198)
Protein-ligand (N=238)
Protein-RNA (N=177)
Protein-DNA (N=254)


d
HF-S1 over HF3
Relative improvement (%)
High
Medium
Low
Probability Groups
Protein-antibody
Protein-protein
Protein-ligand
Protein-RNA
Protein-DNA


e
Association between contact probability and residue interaction types
ALA ARG ASN ASP CYS GLN GLU GLY HIS ILE LEU LYS MET PHE PRO SER THR TRP TYR VAL
Top 10
Bottom 10
Interaction Proportion (%)
10.0%
10.0%
80.0%
30.0%
10.0%
60.0%
Top 10
Bottom 10
Hydrophobic
Electrostatic
Polar
Others


a
Protein-antibody
Protein-protein
Protein-ligand
Protein-RNA
Protein-DNA
HF-S1 DockQ STD
HF3 DockQ STD
HF-S1 Ligand RMSD STD
HF3 Ligand RMSD STD
HF-S1 iLDDT STD
HF3 iLDDT STD
b
Percentage of cases (%)
48.9%
38.0%
13.1%
37.7%
41.7%
20.1%
56.0%
31.5%
12.5%
59.3%
30.5%
10.2%
61.8%
21.7%
16.5%
HF3 std higher
Comparable
HF-S1 std higher
c
Average precision of sampled structure
Average DockQ
Average % Ligand RMSD < 2Å
Average iLDDT
Sampling num
HF3 on Protein-antibody
HF-S1 on Protein-antibody
HF3 on Protein-protein
HF-S1 on Protein-protein
HF3 on Protein-ligand
HF-S1 on Protein-ligand
HF3 on Protein-RNA
HF-S1 on Protein-RNA
HF3 on Protein-DNA
HF-S1 on Protein-DNA
d
HF-S1
HF3
DockQ
Ranking Confidence
Experimental
HF-S1 predictions
HF3 predictions
PDB ID: 8PE1
iLDDT
PDB ID: 8HHM