Dissecting the Black Box of AlphaFold in Protein–Protein Complex Assembly

Shuangjun Li, Zichun Mu, Chengfei Yan*

School of Physics, Huazhong University of Science and Technology, China

Correspondence: chengfeiyan@hust.edu.cn

Abstract

AlphaFold achieves unprecedented accuracy in modeling protein–protein complexes, yet the principles governing complex assembly remain unclear. Here, we develop a unified interpretability framework for AlphaFold-Multimer and AlphaFold3 to dissect the mechanisms underlying complex formation. We demonstrate that inter-protein coevolution is not a major determinant of assembly. Instead, complex structures are primarily driven by monomer geometry together with interface-level pattern matching between backbone complementarity and residue identities. By visualizing the iterative propagation of distance constraints during inference, we uncover a hierarchical process in which monomer-level constraints are established prior to cross-chain interactions, directly demonstrating that inter-chain geometry is inferred from monomer geometries rather than being encoded by coevolutionary signals. Application to antigen–antibody complexes further reveals that reduced prediction accuracy arises from the non-canonical and structurally plastic nature of immune interfaces, identifying accurate modeling of interface conformations and recognition of atypical antigen–antibody interaction patterns as key bottlenecks for improving immune complex prediction.

Introduction

The accurate prediction of protein–protein complex structures is fundamental to understanding cellular organization and molecular function. Recent advances in deep learning, particularly the AlphaFold family of models[1–3], including AlphaFold-Multimer (AFM)[2] and AlphaFold3 (AF3)[3], have achieved unprecedented accuracy in modeling protein–protein complex structures. Despite this success, the mechanistic principles by which these models infer assemble multi-chain complexes remain poorly understood. This lack of interpretability limits our ability to assess model reliability and to systematically improve prediction performance.

A prevailing hypothesis attributes the success of AlphaFold in complex prediction to inter-protein coevolutionary signals encoded in multiple sequence alignments (MSAs)[2–11]. This view is largely motivated by the well-established reliance of monomer structure prediction on intra-protein coevolution, where correlated mutations provide strong evolutionary constraints on residue contacts[1,3,4,12]. By analogy, it has been proposed that inter-protein coevolution could similarly guide the inference of inter-chain interactions[13–19]. However, in contrast to intra-protein signals, inter-protein coevolution is often substantially weaker or entirely absent, particularly in transient interactions and immune recognition, where AlphaFold nonetheless frequently produces plausible complex structures[8,10,20–24]. This apparent discrepancy raises a fundamental question: how are inter-chain geometries inferred in the absence of strong inter-protein coevolutionary coupling?

Here, we develop a unified mechanistic framework to dissect protein complex assembly in AFM and AF3. Through systematic perturbation and controlled alignment analyses, we show that complex assembly is primarily governed by intra-chain structural constraints that define monomer geometry, with inter-chain organization emerging from geometric compatibility and interface sequence pattern matching, while direct inter-protein coevolution plays a minimal role.

To further resolve the internal mechanism of this process, we introduce an interpretability strategy that tracks the propagation of geometric constraints during iterative refinement. By visualizing intra- and inter-chain distance constraints across layers of the Evoformer stack as well as across recycling steps, we reveal a hierarchical formation process in which monomer-level constraints precede cross-chain interactions, directly demonstrating that inter-chain contacts emerge through the progressive propagation of monomer structural constraints rather than being directly encoded by inter-protein coevolution.

Leveraging this framework, we further investigate antigen–antibody complexes, where prediction performance remains comparatively limited[3,8,20,24,25]. We show that reduced accuracy in these systems arises from the non-canonical and structurally plastic nature of immune interfaces, rather than from missing coevolutionary information.

Results

1. Overview of the modeling frameworks of AFM and AF3

We begin by outlining the modeling pipelines of AFM and AF3 for protein complex structure prediction. Both frameworks take amino acid sequences and their corresponding MSAs as inputs, optionally incorporating monomeric structural templates. Following input embedding, representations are processed by the Evoformer in AFM or the Pairformer in AF3, where intra- and inter-protein geometric relationships are jointly modeled. These representations are then passed to the structure generation module, which is built on Invariant Point Attention (IPA) in AFM and a diffusion-based generative model in AF3, to generate the three-dimensional structure of the protein complex[2,3] (Figure 1a).

Both models employ phylogeny-aware MSA pairing strategies[26] intended to capture putative inter-protein coevolutionary signals that may inform interface formation. However, the contribution of inter-protein coevolution to accurate complex structure prediction, whether explicitly encoded through paired MSAs or implicitly embedded in

unpaired MSAs, remains unclear and warrants systematic investigation.

2. Impact of inter-protein coevolutionary information on AFM and AF3 predictions

2.1 Effect of paired MSA-derived inter-protein coevolution on prediction accuracy

To rigorously assess the contribution of paired MSAs to protein complex structure prediction, we constructed a strictly time-segregated benchmark comprising Protein Data Bank (PDB)[27] entries deposited between January 1, 2022 and December 20, 2024, excluding all structures present in the training sets of AFM and AF3 (cutoff: September 30, 2021). After applying resolution and redundancy filters, the final dataset included 200 homodimers and 316 heterodimers. Homodimers were analyzed separately, as their paired MSAs are generated by concatenating identical sequences without cross-chain pairing, avoiding confounding effects from erroneous pairings.

Using ColabFold[28], we first generated the native AFM Paired MSAs[2] for each protein pair, which consist of (i) paired sequences derived from phylogenetically matched homologs and (ii) unpaired sequences organized in a block-diagonal format to preserve alignment consistency. For homomeric targets, this construction naturally reduces to the concatenation of identical monomeric sequences, thereby eliminating the block-diagonal component.

To isolate the effect of sequence pairing, we constructed two controlled variants from the same underlying MSAs: Block MSA, in which all non-query sequences were gap-padded to eliminate inter-chain pairing, and Randomly Paired MSA, in which sequences were randomly paired across chains. These three conditions form a controlled framework for disentangling the role of paired MSAs (Figure 1b). Each alignment was input to AFM and AF3 without templates, and performance was evaluated using DockQ[29] across both homo- and heteromeric complexes.

As shown in Figures 1c-d, model performance measured by mean DockQ shows minimal sensitivity to the type of MSA. Predictions from Block MSAs and native AFM

MSAs are nearly indistinguishable ($|\Delta| \leq 0.2$), indicating that explicit sequence pairing contributes little to predictive accuracy. Randomly Paired MSAs yield only a marginal decrease ($\Delta \in [-0.03, 0]$), likely due to noise introduced by incorrect pairings. To exclude potential bias from training-set similarity, we further evaluated a subset of complexes with low structural similarity to the AFM and AF3 training data (Supplementary Figure 1), thereby minimizing reliance on template-based recognition. Performance on this subset remains consistent with that on the full dataset, with no significant differences across the three MSA strategies. Collectively, these results indicate that inter-protein coevolutionary signals encoded in paired MSAs are not a major determinant of complex structure prediction accuracy in either AFM or AF3.

Although introducing paired MSAs does not result in a significant difference in overall prediction performance (mean DockQ), a case-by-case analysis of predictions from AFM MSA and Block MSA reveals that paired MSAs can indeed improve prediction accuracy in some individual cases (Supplementary Figure 2). However, there are also instances where the introduction of paired MSAs leads to a decrease in prediction accuracy. Therefore, the limited benefits of AFM-generated paired MSAs may stem, in part, from imperfections in the pairing strategy itself. For homomeric interactions, AFM concatenates homologous sequences irrespective of whether they share the same oligomerization mode, such that low-similarity sequences may introduce noise. For heteromeric interactions, AFM's phylogeny-guided pairing does not guarantee that concatenated sequences correspond to bona fide interacting partners.

Notably, in the cases where paired MSAs significantly improve prediction accuracy, the overlap between the AFM and AF3 model predictions is only 20–30% (Supplementary Figure 2). This raises the question of whether the observed improvement is driven by the coevolutionary information encoded in the paired MSAs, or if it results from the inherent randomness and biases of the models. If incorrect pairings are minimized, can paired MSAs truly enhance the overall performance of the models?

To better assess the ideal contribution of paired MSAs to prediction performance and minimize the impact of incorrect pairings, we applied additional stringency filters. For homodimers, only paired sequences with more than 70% sequence identity to the target were retained, while lower-identity homologs were reassigned to the unpaired MSA. For heterodimers, only sequence pairs sharing the same UniProt[30] ID were preserved, while all other sequences were placed in the unpaired component. Sequence pairs with identical UniProt IDs unambiguously originate from the same protein and typically correspond to distinct domains; such pairs therefore represent correct pairings that do not introduce noise from non-interacting proteins. After filtering, we retained interactions containing more than 50 paired sequences, yielding 61 homomeric and 54 heteromeric complexes for subsequent modeling with AFM and AF3. Despite this stringent selection, no significant performance improvement was observed (Figures 1e–f), reinforcing the conclusion that inter-protein coevolutionary signals in paired MSAs are not a primary driver of AFM or AF3 performance in protein complex structure prediction.

2.2 Effect of latent inter-protein coevolution in unpaired MSAs on prediction accuracy

The lack of performance gains from paired MSAs does not, by itself, exclude a role for inter-protein coevolutionary signals in complex structure prediction. One possibility is that, during training, the models may have implicitly learned interaction patterns across large collections of unpaired sequences, thereby acquiring the capacity to extract inter-protein coevolutionary information even in the absence of explicit sequence pairing.

To directly test whether AFM and AF3 rely on inter-protein coevolutionary signals potentially embedded within unpaired MSAs when predicting complex structures, we sought to eliminate any such signals to evaluate the resulting impact on model performance. This task is nontrivial, as AFM constructs MSAs from multiple metagenomic databases that contain numerous sequences lacking reliable species annotations, making it difficult to ensure that unpaired MSAs do not include sequences

derived from the same organism and thus retain latent species-level coevolutionary information.

To overcome this limitation, we regenerated MSAs for proteins in the heterodimer dataset using JACKHMMER[31] searches against the UniRef100[32] database, retaining only sequences with explicit species annotations. These monomeric MSAs were concatenated in block-diagonal format to generate Block MSAs. To further eliminate any inter-protein coevolutionary signal embedded in Block MSAs, we randomly selected two non-overlapping sets of species for the two monomer MSAs and retained only sequences originating from their respective species sets. This procedure ensured that the two MSAs shared no common species, thereby removing potential species-level interaction signals. Block-format interaction MSAs were then reconstructed using the filtered sequences (Figure 1g). These two sets of Block MSAs were subsequently provided as input to AFM and AF3 for heterodimeric complex structure prediction. Model performance showed no significant difference between the two conditions (Figure 1h), indicating that latent inter-protein coevolutionary signals potentially embedded in unpaired MSAs are not a major contributor to predictive accuracy.

Taken together, these results demonstrate that inter-protein coevolutionary information, whether explicitly encoded through paired MSAs or implicitly embedded within unpaired MSAs, is not the primary determinant of AFM or AF3 performance in protein complex structure prediction.

3. Mechanism of complex prediction without inter-protein coevolution

3.1 A hypothesis: Inter-protein spatial constraints derived from monomer geometries

Our finding that inter-protein coevolution is not the primary source of information used by AFM and AF3 raises a central question: how do these models nonetheless achieve high accuracy in protein complex prediction?

A theoretical framework proposed by Roney *et al.*[33] for protein monomer structure

prediction posits that AlphaFold2 exploits coevolutionary signals encoded in MSAs to guide the model toward regions near the global minimum of the free-energy landscape, followed by local refinement to reach the lowest-energy state, with model confidence metrics reflecting this underlying landscape (Figure 2a).

Here, we extend this framework to protein complexes to describe how information is integrated across interacting chains. In AFM and AF3, the representation-learning modules (Evoformer and Pairformer) transform MSA-derived signals into geometric constraints that restrict the conformational search space, guiding the model toward regions near the free-energy minimum. The structure module then refines these candidates toward the lowest-energy state, while confidence metrics continue to reflect the underlying energy landscape (Figure 2b).

This perspective motivates a closer examination of the nature and origin of the geometric constraints governing protein–protein interactions. These constraints can be decomposed into intra-monomer and inter-protein components. While intra-monomer constraints are largely informed by coevolutionary signals within monomer MSAs, the origin of inter-protein constraints remains unclear. Our results indicate that inter-protein coevolution plays a negligible role in complex prediction, suggesting that these constraints are not directly encoded by such signals. Instead, given that binding interfaces are fundamentally determined by the three-dimensional architectures of the individual subunits, we propose that inter-protein spatial constraints are primarily inferred from intrinsic monomer geometries. Specifically, given a protein complex MSA (either paired or block-formatted), the model first infers intra-chain geometric constraints from the monomer sequence information it contains; on this basis, it derives initial inter-chain constraints. These are then iteratively refined through multiple Evoformer/Pairformer layers and recycling rounds, enabling the joint optimization of intra- and inter-chain geometries and ultimately yielding a self-consistent representation for complex structure modeling (Figure 2c).

3.2 Interrogating the hypothesis through template-driven complex prediction

Building on our model that inter-protein spatial constraints are primarily inferred from monomer geometry rather than direct coevolution, we reasoned that monomer templates explicitly encoding accurate monomer geometry should be sufficient to support complex structure prediction (Figure 3a). Although AFM and AF3 allow the use of monomeric templates, these are generally treated as auxiliary to MSAs, and their standalone contribution remains unclear.

To test this, we evaluated AFM and AF3 using only target sequences and monomer templates. We considered three template types: bound-state monomers derived from AF3 predictions under Block MSA conditions, experimentally determined bound-state monomers, and unbound-state monomers predicted independently by AF3. Performance was assessed on curated homo- and heterodimer datasets and compared to predictions using Block MSAs.

Consistent with our hypothesis, predicted bound-state templates achieved accuracy comparable to MSA-based predictions, indicating that monomer geometry captures much of the structural signal for complex assembly typically attributed to evolutionary information. Notably, experimentally determined bound-state templates even yielded substantially higher accuracy, and incorporating MSAs in addition to such templates did not further improve performance (Figure 3b–c; Supplementary Figure 3). These results demonstrate that accurate protein complex structures can be inferred in the absence of MSAs when high-quality monomer geometry is available. However, predicted unbound-state templates lead to a marked reduction in accuracy, highlighting a clear difference between bound- and unbound-state templates that substantially impacts model performance.

To identify the structural determinants of this performance difference, we compared predicted bound and unbound monomer templates to experimentally determined bound conformations using TM-score[34] at the full-chain, interface, and non-interface regions. Predicted bound monomer templates more closely resemble experimental

conformations, with improvements largely confined to interface regions, while non-interface regions show no significant differences (Figure 3d). Further analysis of AF3 predictions using both bound-state and unbound-state templates revealed that successful complex predictions (DockQ ≥ 0.23) are associated with significantly higher template accuracy at the interface, whereas no such difference is observed in non-interface regions (Figure 3e). Consistent with this observation, interface template accuracy shows a significant correlation (Pearson r = 0.575) with complex prediction performance (Supplementary Figure 4).

Collectively, these results indicate that the accuracy of interface regions in monomer templates is a key determinant of complex structure prediction performance, implying that the model may leverage geometric complementarity at protein interfaces to infer inter-chain constraints and guide complex assembly. Notably, interface regions of bound-state monomer structures predicted by AF3 using Block MSA as input exhibit substantially higher accuracy than those of unbound-state monomers predicted using monomer MSAs. This difference propagates to complex structure prediction when these structures are used as monomer templates by AFM/AF3, resulting in deviations in both the global binding mode (Figure 3b–c) and the local interface geometry (Supplementary Figure 5) of the predicted complexes. Together, these observations also suggest that MSA-based modeling of protein–protein complexes allows the model to partially capture interface rearrangements induced by protein–protein interactions, whereas template-only inference has limited ability to correct such discrepancies when the interface geometry deviates from the experimentally observed conformation. Consequently, monomer templates cannot fully replace MSAs in practical applications.

3.3 Effects of interface and non-interface mutations on complex structure prediction

The TM-score-based analysis above primarily evaluates backbone conformations and does not explicitly capture the contribution of side chains—the sequence itself—to protein–protein complex prediction. Consequently, it remains unclear whether interface recognition is governed solely by backbone geometric constraints or also depends on

side-chain identity.

To systematically assess the role of side chains, we introduced random mutations separately at interface and non-interface residues in both homomeric and heteromeric protein–protein interaction datasets, with mutations limited to ≤10% of residues per chain. The resulting mutants were evaluated using two complementary prediction strategies: (i) Block MSA-based prediction, in which MSAs were reconstructed from the mutated sequences, and (ii) monomer template-based prediction, in which structural templates were remodeled to reflect the mutations based on experimentally determined bound-state conformations. By comparing the predicted complexes with their corresponding experimental structures, we quantified the impact of side-chain perturbations on complex assembly under distinct structural and evolutionary information regimes.

Strikingly, mutations at interface residues led to a near-complete loss of predictive accuracy under both settings, whereas mutations at non-interface residues had only a moderate effect (Figure 4a–d). These results demonstrate that side-chain identity at protein–protein interfaces is essential for accurate identification of interaction interfaces and prediction of complex structures, extending beyond backbone geometric complementarity.

3.4 Monomer-derived features drive complex modeling

In summary, our results indicate that AFM and AF3 primarily rely on structural geometric features (inferred from MSAs or monomer templates) and monomer sequences, rather than inter-chain coevolutionary information, to identify compatible interaction interfaces for protein complex prediction (Figure 4e). This supports a mechanism in which monomer-level geometric constraints are first established by representation-learning modules (Evoformer or Pairformer), followed by the inference of inter-chain constraints, which are then iteratively refined—together with intra-chain constraints—across multiple layers and recycling rounds, enabling their joint optimization and yielding a self-consistent representation for complex structure

modeling.

This mechanism also explains our observation that MSA-based approaches more effectively capture interaction-induced interfacial conformational changes than template-based approaches. In MSA-based inference, intra-chain constraints are progressively constructed and refined during model inference, enabling inter-chain constraints to be jointly updated in a coupled manner and thereby supporting coordinated folding and docking. In contrast, template-based inputs largely fix intra-chain geometry at an early stage by directly inheriting template-derived structures, after which inter-chain interactions are inferred primarily through geometric complementarity and sequence compatibility, resembling a more rigid docking process and limiting conformational adaptation upon binding.

It is worth noting that, in cases where strong inter-protein coevolutionary signals are present, paired MSAs may still contribute to improved prediction performance. However, our comprehensive analyses indicate that, for the vast majority of protein complexes, inter-protein coevolution is not the primary determinant of predictive accuracy. Instead, successful complex structure prediction is governed by accurate inference of monomeric geometry and interface-specific conformational adaptability.

4. Validating the mechanism by tracking spatial constraints in AFM modeling

To further validate the mechanistic basis of AlphaFold-based protein complex prediction, we introduce a method termed AlphaFold-Constraint Propagation Mapping (AF-CPM) for visualizing the evolution of intra- and inter-chain geometric constraints during AFM inference. AF-CPM leverages the distogram head implemented in OpenFold[35] to transform pair representations from successive Evoformer layers into residue–residue distance distributions. By aggregating probabilities corresponding to distances below 12 Å, we obtain residue–residue contact probability maps (Figure 5a). This framework enables direct visualization of how spatial geometric constraints are progressively inferred and propagated during complex structure prediction, providing a mechanistically interpretable view of the model's internal inference process.

We applied AF-CPM to a representative heterodimeric complex (PDB ID: 8G4K) (Figure 5b). When the model was provided with protein sequences alone, without MSAs, it failed to infer intra-chain contact maps and the monomeric structure of one interacting partner (chain B), and was also unable to recover the correct inter-chain interaction interface (Figure 5c, 5e; Supplementary Figure 6). This confirms that coevolutionary information encoded in MSAs is important for accurate monomer structure inference. In contrast, when MSAs generated by AFM—including both paired and block-format alignments—were supplied, the model first established robust intra-chain contact maps and subsequently inferred inter-chain contact patterns (Figure 5f; Supplementary Figure 7). Notably, inter-chain contacts emerged only after intra-chain constraints had been established, indicating that inter-protein geometric relationships are inferred downstream of monomer-level geometry rather than being directly extracted from inter-protein coevolutionary signals.

When MSAs were omitted and experimentally determined bound-state monomer structures were provided as templates, the model accurately recovered both intra- and inter-chain contact maps and produced correct complex structures. In this template-based setting, intra-chain contact information was directly inherited from the template, bypassing the need for progressive construction through Evoformer layers, while inter-chain constraints were inferred based on the given monomeric geometry (Figure 5g; Supplementary Figure 8). This observation suggests that monomer-template–based modeling does not involve a progressive process for establishing intra-chain geometric constraints, which may explain its limited ability to capture interaction-induced conformational rearrangements at protein–protein interfaces.

To further interrogate the role of interface residues in complex assembly, we introduced random mutations at interface residues of one interaction partner (chain A) and remodeled the mutated sequence using the original monomeric structural template. Despite the preservation of strong intra-chain contact signals, inter-chain contact

probabilities were almost completely abolished (Figure 5h; Supplementary Figure 9). Consequently, the model failed to recover the original binding mode, even though the monomeric structure itself remained accurately predicted (Figure 5d). These results demonstrate that accurate complex assembly cannot be achieved solely from monomer backbone geometry and highlight a critical dependence of inter-chain contact inference on interface residue identity. Together, these findings further validate the proposed protein–protein assembly mechanism in AlphaFold, whereby monomer-level geometric constraints are first established and subsequently leveraged to identify binding modes through the joint matching of interface backbone geometry and residue identities.

5. Origins of limited prediction accuracy in antigen–antibody systems

Although AlphaFold models have demonstrated remarkable performance in modeling conventional protein–protein complexes, their performance on antigen–antibody systems remains substantially lower[24,25]. This limitation has often been attributed to the lack of inter-chain coevolutionary signals at the interface. However, our prior analyses of AlphaFold's complex assembly mechanism indicate that the model primarily relies on geometric complementarity and sequence compatibility at the interface, rather than inter-chain coevolution. This raises the question of whether the absence of inter-chain coevolutionary signals is indeed the primary factor limiting performance. Motivated by this, we constructed an independent benchmark of 197 nonredundant antigen–antibody complexes to systematically reassess this assumption.

We first evaluated AFM and AF3 under multiple input conditions, including AFM MSAs, Block MSAs, and bound-state monomer templates derived from experimental structures. Across all settings, introducing inter-chain sequence pairing did not improve prediction accuracy for either antigen–antibody complexes or antibody heavy–light chain assembly. In contrast, bound-state monomer templates consistently outperformed all MSA-based approaches (Figure 6a–b). This trend is consistent with observations in conventional homo- and heterodimer systems, supporting that the reduced performance is not attributable to the absence of inter-chain coevolution, but instead reflects a geometry-driven assembly process.

Nevertheless, this mechanistic insight does not fully account for the reduced prediction accuracy observed for antigen–antibody complexes. To investigate the origin of this discrepancy, we systematically compared monomer-level structural accuracy between antigen–antibody and conventional heterodimer datasets using AF3-predicted complex structures with block MSAs as input. Performance was evaluated across the full-chain, interface, and non-interface regions.

We found that antibody monomers (with the heavy and light chains merged for evaluation) show comparable full-chain and non-interface region accuracy to those of control heterodimers, whereas antigen monomer accuracy is significantly lower across full-chain, interface, and non-interface regions. More importantly, local structural accuracy at the interface—defined based on experimentally determined complex structures—on both the antigen and antibody sides is substantially lower than that of conventional heterodimers (Figure 6c). This reduced interfacial accuracy correlates with overall complex accuracy, with the most pronounced effects observed in the antibody complementarity-determining region (CDR)-H3 loop (Figure 6d; Supplementary Figure 10).

Together, these results identify a central bottleneck: the limited ability of AlphaFold to model structurally plastic interface regions. Antibody specificity is largely determined by somatically hypermutated CDRs, particularly the highly variable CDR-H3 loop, while antigen epitopes evolve under immune selection and display substantial sequence and conformational diversity. As a result, these rapidly evolving and conformationally heterogeneous interfaces constitute intrinsically challenging targets in monomeric structure prediction, thereby impairing the model's ability to infer correct binding geometries through geometric matching.

Further statistical analysis of interfacial residue contacts revealed systematic differences between antigen–antibody complexes and conventional heterodimers. The

antigen–antibody interface exhibits a distinctive residue usage pattern, with pronounced enrichment of tyrosine and serine on the antibody side—features not characteristic of typical protein–protein interfaces (Figure 6e). In addition, both heterodimer and homodimer datasets show clear differences in contact frequency distributions between successfully and unsuccessfully predicted cases (Supplementary Figure 11), highlighting a key limitation in the model's learned interface priors and suggesting that prediction performance is governed by the alignment between interface features and learned statistical distributions. Since the model is predominantly trained on non–antigen–antibody complexes, these priors are better aligned with conventional protein–protein interaction statistics, leading to a distribution mismatch when applied to immune complexes, which may contribute as an additional source of reduced prediction performance on antigen–antibody targets.

Finally, we examined the evolution of geometric constraints in AFM predictions of antigen–antibody complexes using the AF-CPM method proposed in this study. Using PDB ID: 8BK2 as an example, AFM with its default MSA input required at least five recycling iterations to successfully assemble the complex (Figure 6f-g). In the early recycling steps, the model first establishes intra-chain constraints for each monomer, including the antibody's heavy and light chains, as well as the antigen. It then forms inter-chain constraints between the antibody's heavy and light chains, assembling the antibody structure correctly. However, at the end of the first recycling round, no constraints exist between the antigen and antibody. Only after five iterations of recycling do cross-chain constraints between the antibody and antigen gradually form, leading to the convergence of the complex into its correct conformation. (Figure 6g; Supplementary Figures 12–13). Notably, successful convergence coincides with improved local accuracy at the antibody interface, indicating a tight coupling between local interface refinement and global complex assembly.

Together, these findings demonstrate that the primary limitation in antigen–antibody prediction lies not in the absence of inter-chain coevolution, but in the structurally

plastic, rapidly evolving, and statistically atypical nature of immune interfaces. This highlights local interfacial conformational accuracy and atypical interface statistics as key factors underlying AlphaFold's limited performance on these targets.

Conclusion

In this work, we establish a mechanistic framework for protein complex structure prediction in AlphaFold-family models. Through systematic perturbation and controlled alignment analyses, we show that inter-protein coevolution is not the dominant determinant of prediction accuracy. Instead, complex assembly is driven predominantly by monomeric geometric architecture, with inter-chain organization emerging through interface-level geometric and sequence pattern matching.

Our analyses, supported by AF-CPM—a method introduced in this study for tracking the propagation of residue-level geometric constraints during AlphaFold inference—reveal a hierarchical inference process in which intra-chain structural constraints are resolved first and subsequently guide the formation of inter-chain interactions. This directly demonstrates that inter-chain structural constraints are progressively established upon intra-chain structural constraints, rather than being directly encoded in inter-protein coevolutionary signals. Consistently, template-driven modeling demonstrates that accurate monomer geometry can substitute for MSA-derived evolutionary information, although MSAs remain more effective in capturing interaction-induced conformational plasticity.

Importantly, we identify structural plasticity and statistical heterogeneity at immune interfaces as principal bottlenecks limiting prediction accuracy for antigen–antibody complexes. Unlike conventional protein–protein interfaces, immune recognition surfaces undergo rapid evolutionary diversification, exhibit distinctive residue composition patterns, and display pronounced local conformational variability—particularly within antibody CDR-H3 loops. These features create a distributional mismatch relative to the structural priors encoded in current AlphaFold models, thereby

reducing the reliability of geometric and sequence-based interface matching.

Collectively, this work challenges the prevailing view that inter-protein coevolution is the primary determinant of deep-learning-based complex prediction. Instead, our results indicate that predictive performance is largely governed by the accuracy of monomer-level structural modeling at interface regions, which directly determines the reliability of interface-specific structural compatibility. This framework provides a mechanistic basis for understanding AlphaFold-based oligomer modeling and suggests that future improvements should prioritize more accurate modeling of structurally heterogeneous and conformationally flexible interface regions, as well as methods capable of robustly handling interaction systems with broad sequence and structural diversity.

Furthermore, while this study primarily focuses on investigating the intrinsic mechanisms of AlphaFold-based modeling of protein oligomers, the analytical approach and the geometric constraint propagation mapping method developed for visualizing the structural prediction process are also widely applicable to other biomolecular interaction systems, including nucleic acids, small molecules, and beyond.

Methods

1. Dataset preparation

1.1 Homodimer dataset

Crystal structures released between 1 January 2022 and 20 December 2024 were retrieved from the PDB. Structures were selected based on two criteria: resolution ≤3.0 Å and biological assemblies composed of two identical protein chains, yielding 7,620 structures. To remove sequence redundancy, the remaining structures were clustered using MMseqs2[36] at a 40% sequence identity threshold, resulting in 2,509 clusters. From these, 200 homodimer complexes were randomly selected to construct a manageable benchmark dataset, each representing a distinct cluster and restricted to entries with a unique biological assembly.

1.2 Heterodimer dataset

Using the same temporal and resolution criteria as for the homodimer dataset, 2,310 PDB entries containing heterodimeric structures were retrieved from the PDB. To remove sequence redundancy, heterodimeric chains were clustered using MMseqs2 at a 40% sequence identity threshold, resulting in 691 clusters. Each heterodimeric chain pair was then assigned a cluster pair identifier based on the cluster membership of its constituent chains, treating the pair as unordered (i.e., chain cluster A–chain cluster B and chain cluster B–chain cluster A were considered equivalent), and pairs in which both chains belonged to the same cluster were excluded to avoid sequence homology between interacting partners. Heterodimer complexes sharing the same cluster pair identifier were grouped, and one representative complex was randomly selected from each group, restricted to entries with a unique biological assembly, yielding 317 non-redundant heterodimer complexes. One covalently linked dimer (PDB ID: 8DLE) was subsequently excluded, resulting in a final benchmark set of 316 heterodimer complexes.

1.3 Antigen–antibody dataset

Antigen–antibody complex structures were retrieved from the Structural Antibody Database[37] (SAbDab) using the "search structures by attribute" function. Structures were filtered with a resolution cutoff of ≤4.0 Å and antigen type restricted to proteins. To avoid overlap with the training data of AFM/AF3, only structures released between 1 January 2022 and 18 March 2024 were included, yielding 2083 antigen–antibody complexes. To remove redundancy, antigen sequences were clustered using MMseqs2 at a 40% sequence identity threshold. Antigen–antibody complexes sharing the same antigen identifier were grouped, and one representative complex was randomly selected from each group, resulting in a final test set of 197 antigen–antibody complexes.

2. Structure prediction implementation

2.1 AFM

Protein complex structure prediction was performed using AlphaFold2_multimer_v3 as implemented in LocalColabFold (https://github.com/YoshitakaMo/localcolabfold). For each target, five models were generated using the five sets of model parameters with a

fixed random seed (0). The model with the highest ranking score was selected as the final prediction.

2.2 AF3

Protein complex structure prediction was performed using the official AlphaFold3 scripts (https://github.com/google-deepmind/alphafold3). For each target, five models were generated with a fixed random seed (1). The model with the highest ranking score was selected as the final prediction.

3. Design and implementation of interface and non-interface mutations

3.1 Selection of mutation sites

Protein–protein interaction interfaces were defined from dimeric complex structures retrieved from the PDB. A residue was classified as an interface residue if any of its heavy atoms lay within 5 Å of any heavy atom from the partner chain, whereas residues with a minimum heavy-atom distance greater than 8 Å from all partner-chain atoms were defined as non-interface residues.

For heterodimers, mutations were introduced randomly in a single chain. For homodimers, to preserve structural symmetry, only residues with consistent interface or non-interface classification at equivalent positions in both chains were considered, and mutations were introduced at these positions in both chains simultaneously. Five homodimers lacking residues that satisfied these criteria were excluded from further analysis.

Interface and non-interface perturbations were performed in separate experiments. For each category, the number of mutated residues was limited to at most 10% of the chain length; when the number of eligible residues exceeded this threshold, a random subset of residues corresponding to 10% of the chain length was selected. All selected residues were mutated to glycine to standardize the perturbation.

3.2 Generation of MSAs for mutated sequences

The mutated sequences were submitted to LocalColabFold to generate MSAs using the AlphaFold2_multimer_v3 model. MSAs were preprocessed to remove lowercase

insertion characters, and the paired segment was extracted and converted into a block MSA format for subsequent complex structure prediction.

3.3 Generation of monomer templates for mutated sequences

Single-chain templates were obtained from experimentally determined bound-state structures in the PDB. Mutated sequences were aligned to these templates, and homology modeling was performed using MODELLER[38] to incorporate mutations while preserving the backbone conformation. The resulting monomer templates were used for complex structure prediction with the mutated sequences.

4. Quantitative assessment of predictions

4.1 Protein–protein complex structure assessment

The quality of protein–protein complex structure predictions was evaluated using DockQ (https://github.com/wallnerlab/DockQ).The prediction quality is categorized based on the DockQ score as follows: 0 ~ 0.23 as incorrect, 0.23 ~ 0.49 as acceptable quality, 0.49 ~ 0.8 as medium quality, and 0.8 ~ 1 as high quality. For mutation experiments, to address sequence–structure mismatches introduced by mutations, predicted structures were remodeled onto the corresponding wild-type sequences using MODELLER prior to DockQ calculation. For antigen–antibody complexes, which involve three chains (antibody heavy and light chains and one antigen chain), DockQ was computed under two evaluation settings. For antigen–antibody interface evaluation, the antibody heavy and light chains were merged into a single chain, whereas for evaluation of the antibody heavy–light chain interface, the two chains were treated as separate entities.

4.2 Monomer structure assessment

The accuracy of monomer structures was assessed using TM-score, computed with USalign[39]. The multimeric alignment option (-mm) was set to 0 to perform pairwise monomer structure alignment. To compare interface and non-interface regions, the interface region was defined as residues within 5 Å of any heavy atom on the partner chain, whereas the non-interface region comprised residues with a minimum distance greater than 8 Å to any heavy atom on the partner chain. For antibody structures, the heavy and light chains were treated as a single entity and evaluated jointly, except when

analyzing CDR loops (Figure 6d). Accordingly, antibody monomer accuracy in Figure 6c and Supplementary Figure 10a was computed based on the combined heavy–light chain structure.

4.3 Statistical analysis

Because DockQ and TM-score distributions deviated from normality, statistical comparisons between groups were performed using the two-sided Mann–Whitney U test[40]. Differences with $p < 0.05$ were considered statistically significant.

5. Implementation of AF-CPM for tracking geometric constraints

We employed OpenFold (v2.2.0) with pretrained AFM parameters (params_model_3_multimer_v3) to track the propagation of residue-level geometric constraints during protein complex prediction. The OpenFold pipeline was modified to support customized template and MSA inputs.

To extract intermediate structural representations, we further modified OpenFold to retrieve pair representations from each of the 48 layers of the Evoformer stack, which were subsequently converted into residue–residue distance distributions using the Distogram head. The resulting distance distributions were normalized using a softmax function, such that each bin represents the predicted probability that the Cβ atoms of a residue pair fall within the corresponding distance interval. Contact probability maps were then derived by summing the probabilities of the first 32 distance bins, corresponding to a distance threshold of 12 Å. The resulting layer-wise contact probability maps capture the progressive establishment of geometric constraints during complex structure prediction.

Code and Data Availability

All data supporting the findings of this study are available within the paper and its Supplementary Information. The PDB IDs of all datasets used, together with the raw data underlying the main results, are available at https://github.com/ChengfeiYan/AF-CPM/tree/main/data. Code for running AF-CPM is available in the same repository. Additional scripts used for data processing and analysis are available from the

corresponding author upon reasonable request.

Acknowledgements

The work was supported by the National Natural Science Foundation of China (grant number: 32571438 and 32101001). Computational resources were provided by the HPC platform of Huazhong University of Science and Technology.

Author Contributions

C.Y. conceived the project, designed the experiments, and supervised the study. S.L. performed the experiments. S.L., M.Z., and C.Y. analyzed the data. S.L. and C.Y. drafted the manuscript. All authors discussed the results and revised the manuscript.

Ethics Declarations

The authors declare no competing interests.

References

1. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
2. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.04.463034 (2022).
3. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
4. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876 (2021).
5. Humphreys, I. *et al.* Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805 (2021).
6. Si, Y. & Yan, C. Protein complex structure prediction powered by multiple sequence alignments of interologs from multiple taxonomic ranks and AlphaFold2. *Brief. Bioinform.* **23**, bbac208 (2022).
7. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein–protein

interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022).

8. Zheng, W. *et al.* Improving deep learning protein monomer and complex structure prediction using DeepMSA2 with huge metagenomics data. *Nat. Methods* **21**, 279–289 (2024).
9. Humphreys, I. R. *et al.* Protein interactions in human pathogens revealed through deep learning. *Nat. Microbiol.* **9**, 2642–2652 (2024).
10. Feldman, J. & Skolnick, J. AF3Complex yields improved structural predictions of protein complexes. *Bioinformatics* **41**, btaf432 (2025).
11. Zhang, J. *et al.* Predicting protein–protein interactions in the human proteome. *Science* **390,** eadt1630 (2025).
12. Ovchinnikov, S. *et al.* Protein structure determination using metagenome sequence data. *Science* **355**, 294-298 (2017).
13. Szurmant, H. & Weigt, M. Inter-residue, inter-protein and inter-family coevolution: bridging the scales. Curr. Opin. Struct. Biol. **50**, 26-32 (2018).
14. Hopf, T. A. *et al.* Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* **3**, e03430 (2014).
15. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *Elife* **3**, e02030 (2014).
16. Gueudré, T., Baldassi, C., Zamparo, M., Weigt, M. & Pagnani, A. Simultaneous identification of specifically interacting paralogs and interprotein contacts by direct coupling analysis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12186-12191(2016).
17. Bitbol, A. F., Dwyer, R. S., Colwell, L. J. & Wingreen, N. S. Inferring interaction partners from protein sequences. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12180-12185 (2016).
18. Cong, Q., Anishchenko, I., Ovchinnikov, S. & Baker, D. Protein interaction networks revealed by proteome coevolution. *Science* **365**, 185-189 (2019).
19. Green, A. G. *et al.* Large-scale discovery of protein interactions at residue resolution using co-evolution calculated from genomic sequences. *Nat. Commun.* **12**, 1396 (2021).

20. Xu, D., Chu, L.-S. & Gray, J. J. Adapting Co-Folding Models for Structure-Based Protein–Protein Docking Through Flow Matching. Preprint at *bioRxiv* https://doi:10.1101/2025.11.28.691195 (2025).

21. Harmalkar, A., Lyskov, S. & Gray, J. J. Reliable protein–protein docking with AlphaFold, Rosetta, and replica exchange. *Elife* **13**, RP94029 (2025).

22. Gao, M., Nakajima An, D., Parks, J. M. & Skolnick, J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat. Commun.* **13**, 1744 (2022).

23. Gao, M. & Skolnick, J. Improved deep learning prediction of antigen–antibody interactions. *Proc. Natl. Acad. Sci.* **121**, e2410529121 (2024).

24. Yin, R. & Pierce, B. G. Evaluation of AlphaFold antibody–antigen modeling with implications for improving predictive accuracy. *Protein Sci.* **33**, e4865 (2024).

25. Hitawala, F. N. & Gray, J. J. What does AlphaFold3 learn about antigen and nanobody docking, and what remains unsolved? Preprint at *bioRxiv* https://10.1101/2024.09.21.614257 (2025).

26. Zeng, H. *et al.* ComplexContact: A web server for inter-protein contact prediction using deep learning. *Nucleic Acids Res.* **46**, W432-W437 (2018)

27. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

28. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

29. Mirabello, C. & Wallner, B. DockQ v2: improved automatic quality measure for protein multimers, nucleic acids, and small molecules. *Bioinformatics* **40**, btae586 (2024).

30. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).

31. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).

32. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).

33. Roney, J. P. & Ovchinnikov, S. State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold. *Phys. Rev. Lett.* **129**, 238101 (2022).

34. Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).

35. Ahdritz, G. *et al.* OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nat. Methods* **21**, 1514–1524 (2024).

36. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

37. Dunbar, J. *et al.* SAbDab: the structural antibody database. *Nucleic Acids Res.* **42**, D1140–D1146 (2013).

38. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinforma.* **Chapter 5**, Unit-5.6 (2006).

39. Zhang, C., Shine, M., Pyle, A. M. & Zhang, Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **19**, 1109–1115 (2022).

40. McKnight, P. E. & Najab, J. Mann-Whitney U Test. in *The Corsini Encyclopedia of Psychology* 1 (John Wiley & Sons, Ltd, 2010).

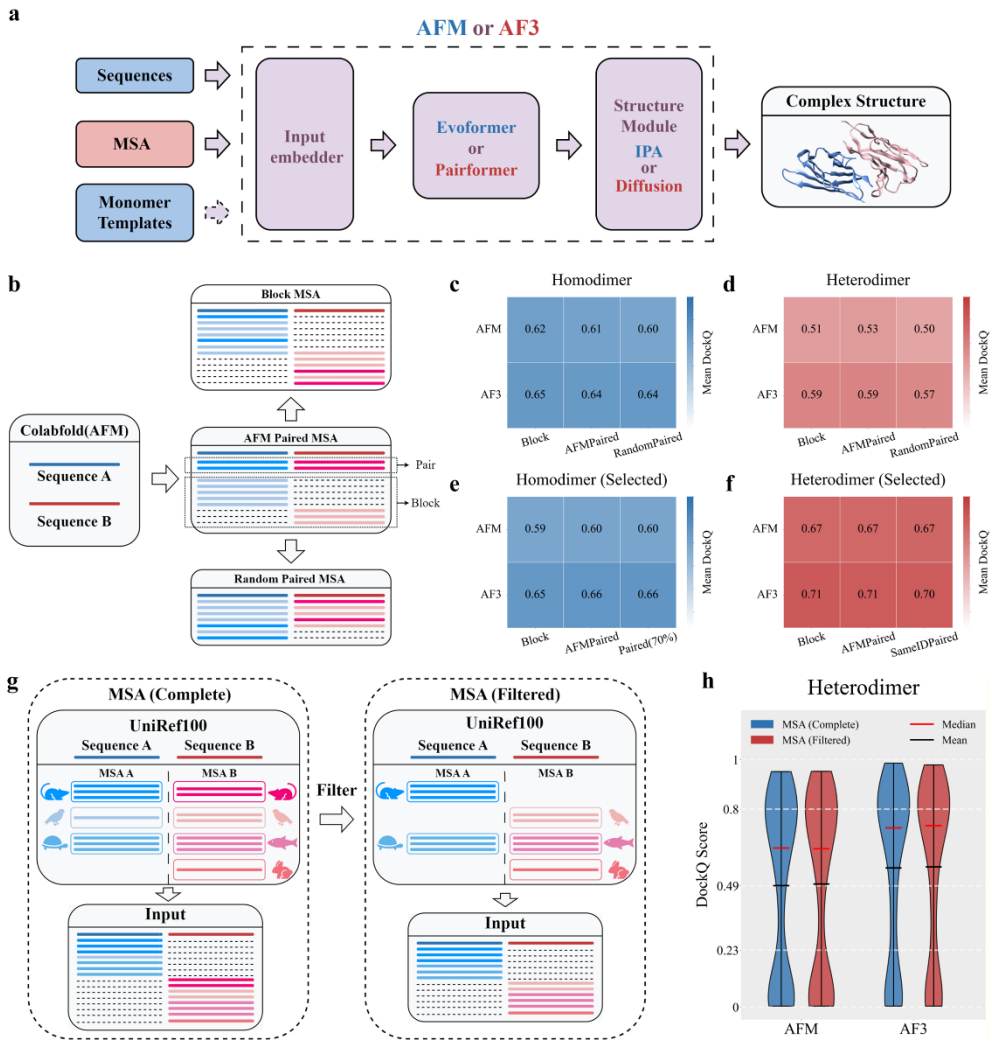


Figure 1. Effects of MSA pairing strategies on protein complex structure prediction.

(a) AFM and AF3 pipelines for protein complex modelling.

(b) Schematic of MSA input strategies: Block MSA, AFM Paired MSA, and Randomly Paired MSA.

(c, d) Mean DockQ scores for homodimer (c) and heterodimer (d) datasets using AFM and AF3 with different MSA inputs.

(e) Performance on the filtered homodimer dataset using Block MSA, AFM Paired MSA , and Block MSA with paired sequences at ≥70% identity to the target.

(f) Performance on the filtered heterodimer dataset using Block MSA alone, AFM Paired MSA, and Block MSA with paired sequences sharing identical UniProt IDs.

(g) Schematic of Block MSA construction in which the MSAs of the two interacting proteins contain no shared species.

(h) Performance of AFM and AF3 with or without species overlap between partner MSAs. The white dashed horizontal line represents key thresholds based on the DockQ evaluation of the predicted results: incorrect (0 ~ 0.23), acceptable quality (0.23 ~ 0.49), medium quality (0.49 ~ 0.8), and high quality (0.8 ~ 1).

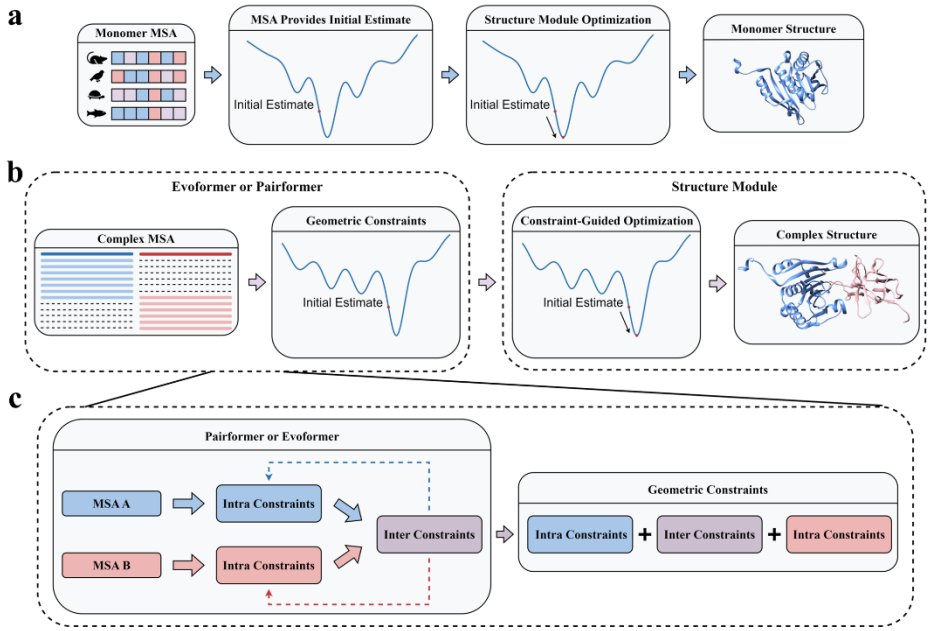


Figure 2. Mechanistic insights into AlphaFold predictions of monomeric and complex structures from MSAs.

(a) AlphaFold generates an initial structural estimate from monomer MSAs, which is refined by the structure module to yield the final monomer structure. Image inspired by Roney et al.[33].

(b) For protein–protein complexes, geometric constraints inferred from complex MSAs through the Evoformer/Pairformer module provide an initial estimate of the complex structure, which is subsequently refined in the structure module to yield the final complex structure.

(c) In the Evoformer/Pairformer module, intra-chain constraints are first extracted from monomer MSAs and used to initialize inter-chain constraints, which are then progressively refined through iterative updates, yielding a self-consistent representation for complex structure prediction.

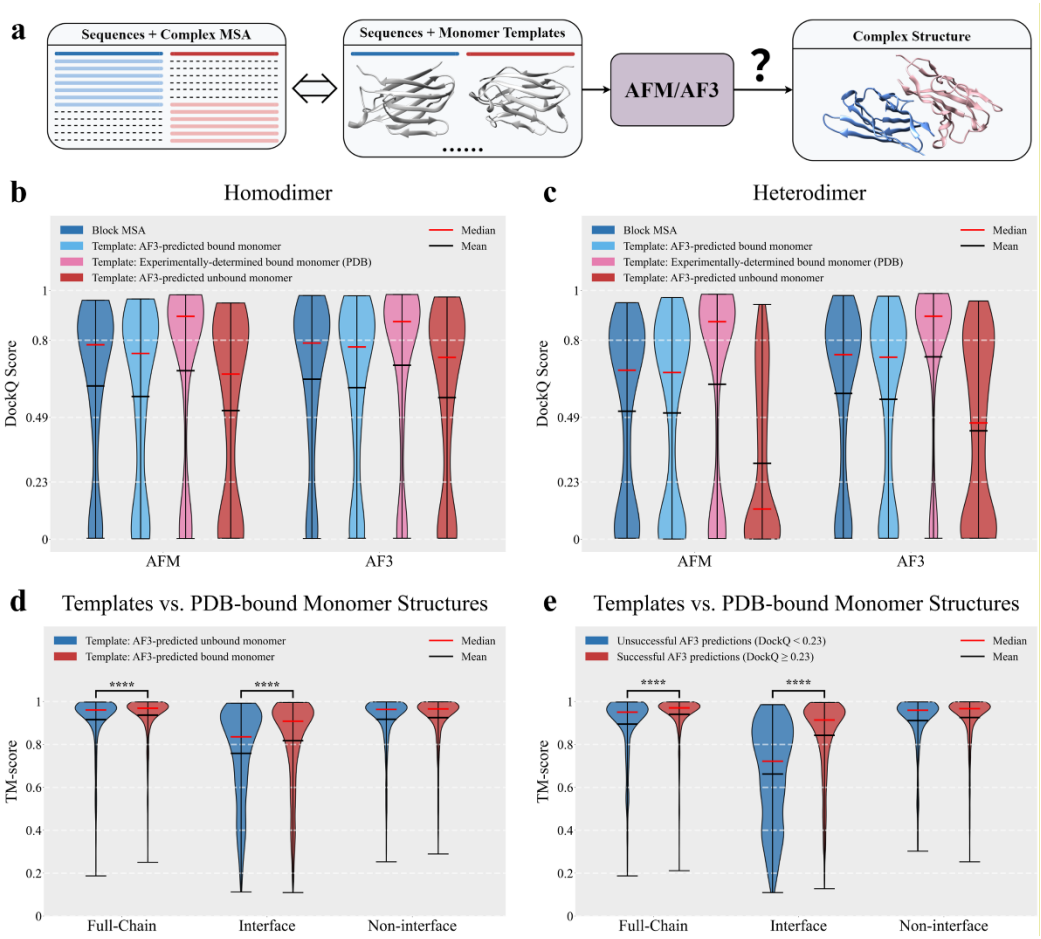


Figure 3. Monomer template-driven protein complex structure prediction.

(a) Use of monomer templates as substitutes for complex MSA in complex prediction.

(b, c) Performance on homodimer (b) and heterodimer (c) datasets using Block MSA versus three types of monomer templates: AF3-predicted bound templates, experimentally determined bound monomer templates (PDB), and AF3-predicted unbound templates.

(d) TM-score distributions of predicted bound and unbound monomer templates relative to experimental bound monomers (PDB), evaluated at the full-chain, interface, and non-interface regions. Significant differences were observed at the full-chain and interface regions (****$p<0.0001$, two-sided Mann–Whitney U test), but not at the non-interface region ($p>0.05$).

(e) TM-score distributions of predicted monomer templates relative to experimental bound monomers (PDB), comparing successful and unsuccessful AF3 predictions at the full-chain, interface, and non-interface regions. Significant differences were observed at the full-chain and interface regions (****$p<0.0001$), but not at the non-interface region ($p>0.05$).

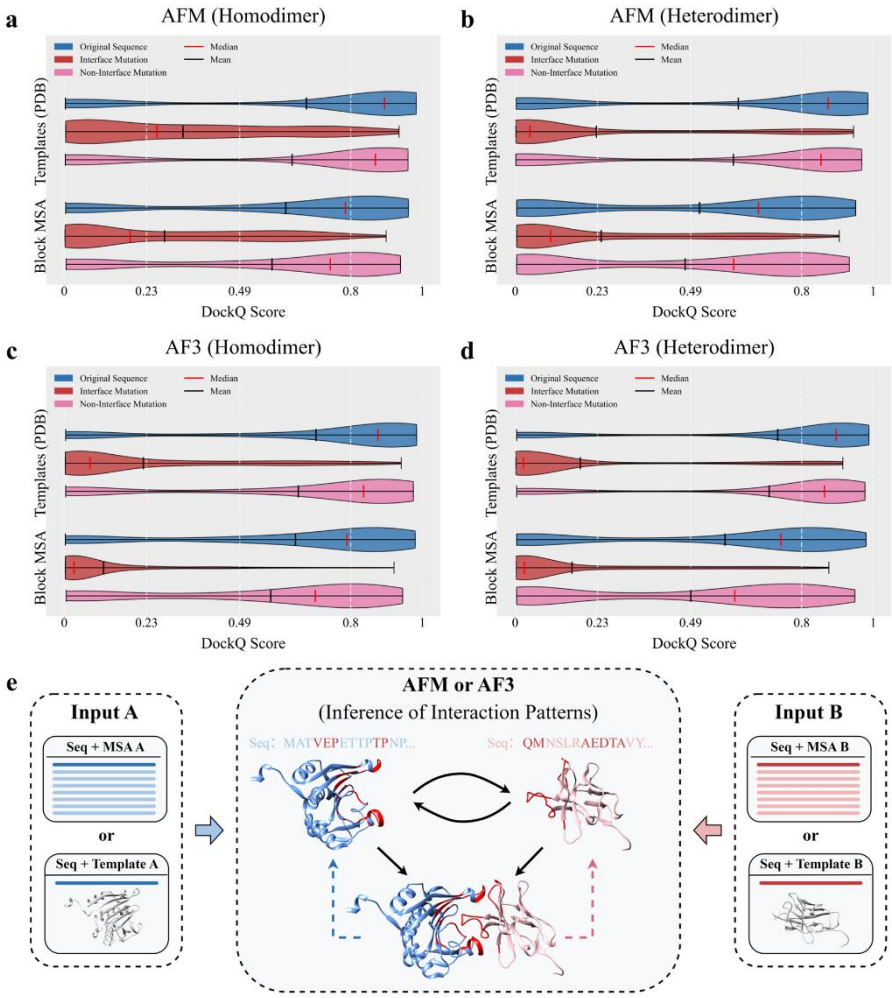


Figure 4. Effects of interface and non-interface mutations on complex structure prediction and the underlying mechanism.

(a–d) AFM (a, b) and AF3 (c, d) performance on homodimer (a, c) and heterodimer (b, d) datasets using Block MSA or experimental determined bound monomer templates (PDB), after introducing independent random mutations at interface or non-interface residues.

(e) **Modeling mechanism:** The model first establishes monomer-level geometric constraints from sequence information and MSAs or templates, and then infers compatible inter-chain interaction patterns based on the resulting monomer sequence features and structural geometry. These inter-chain constraints are subsequently iteratively refined together with intra-chain constraints, yielding a self-consistent representation for complex structure modeling.

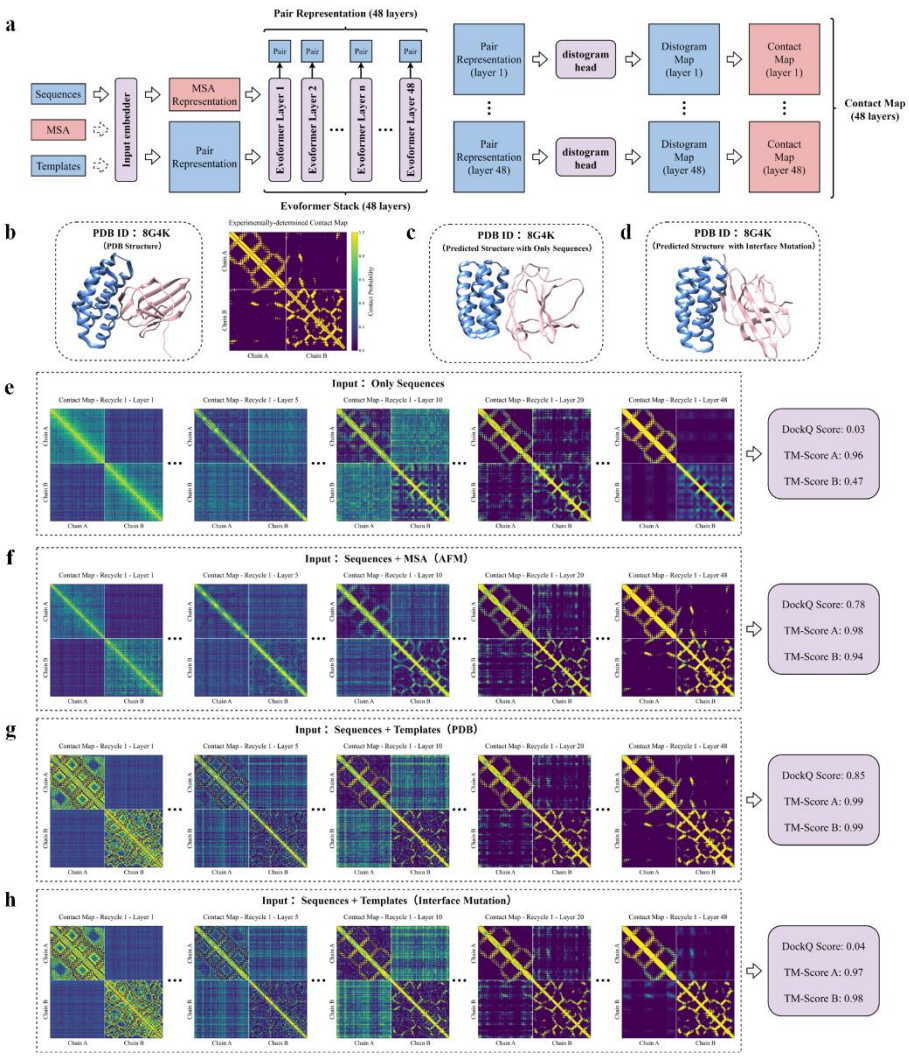


Figure 5. AF-CPM reveals spatial constraint propagation during AFM inference.

(a) **Overview of AF-CPM:** Pair representations from each layer of the 48-layer Evoformer stack are extracted and passed to the distogram head to generate distogram maps. Contact probability maps are obtained by summing the probabilities over the first 32 distance bins of the distograms, corresponding to a distance threshold of 12 Å. These contact probability maps are computed for each layer of the Evoformer stack, yielding layer-wise representations that reflect the progressive formation of geometric constraints during complex structure prediction.

(b) Experimentally determined structure of 8G4K (PDB) and its corresponding contact map derived from the experimental structure.

(c) Predicted complex structure of 8G4K using only sequences.

(d) Predicted complex structure of 8G4K after mutating interface residues on chain A, using mutated monomer templates.

(e–h) Tracking constraint formation using AF-CPM. Evolution of spatial constraints across Evoformer layers and the corresponding prediction quality for 8G4K under different input conditions: sequence only (f), sequence + MSA (AFM) (g), sequence + templates (PDB) (h), and sequence + templates with interface mutations (i).

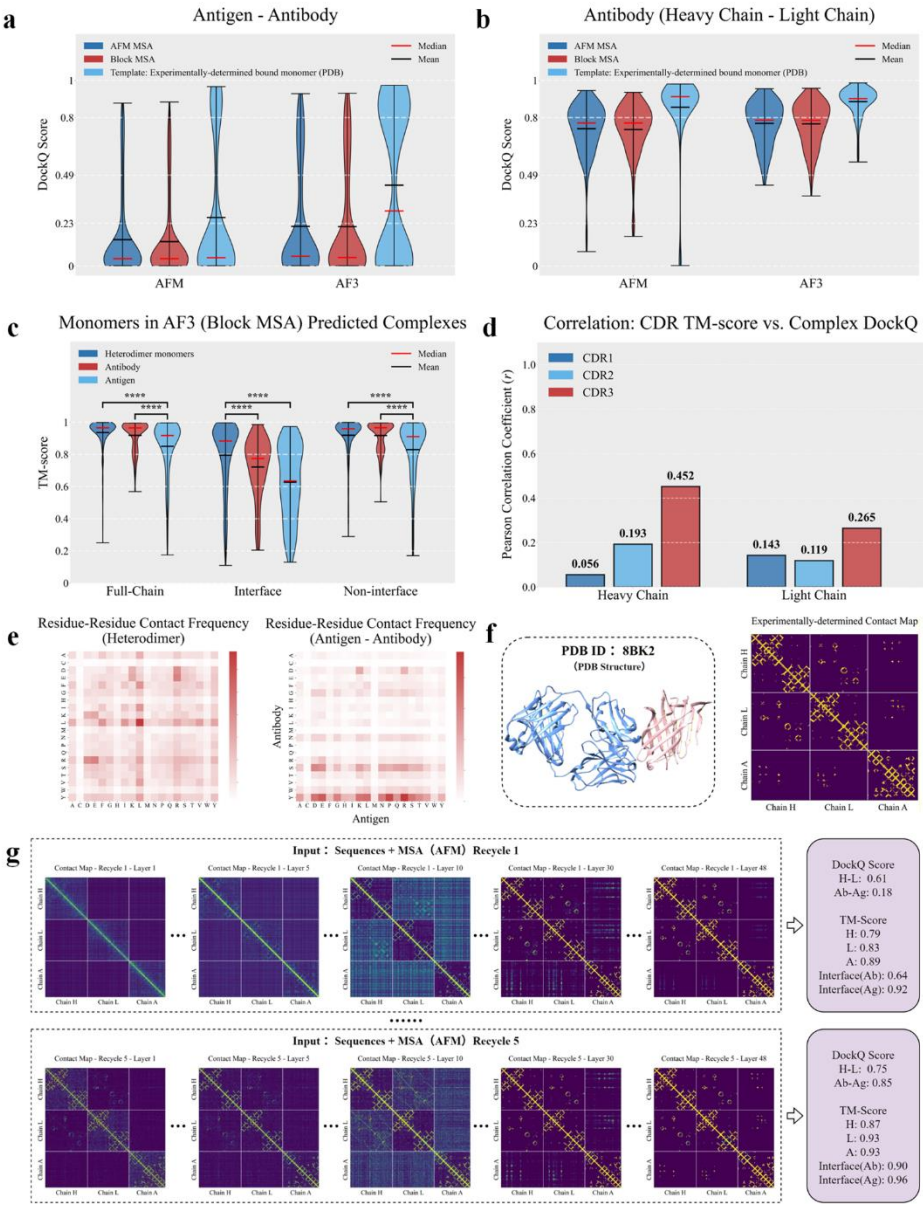
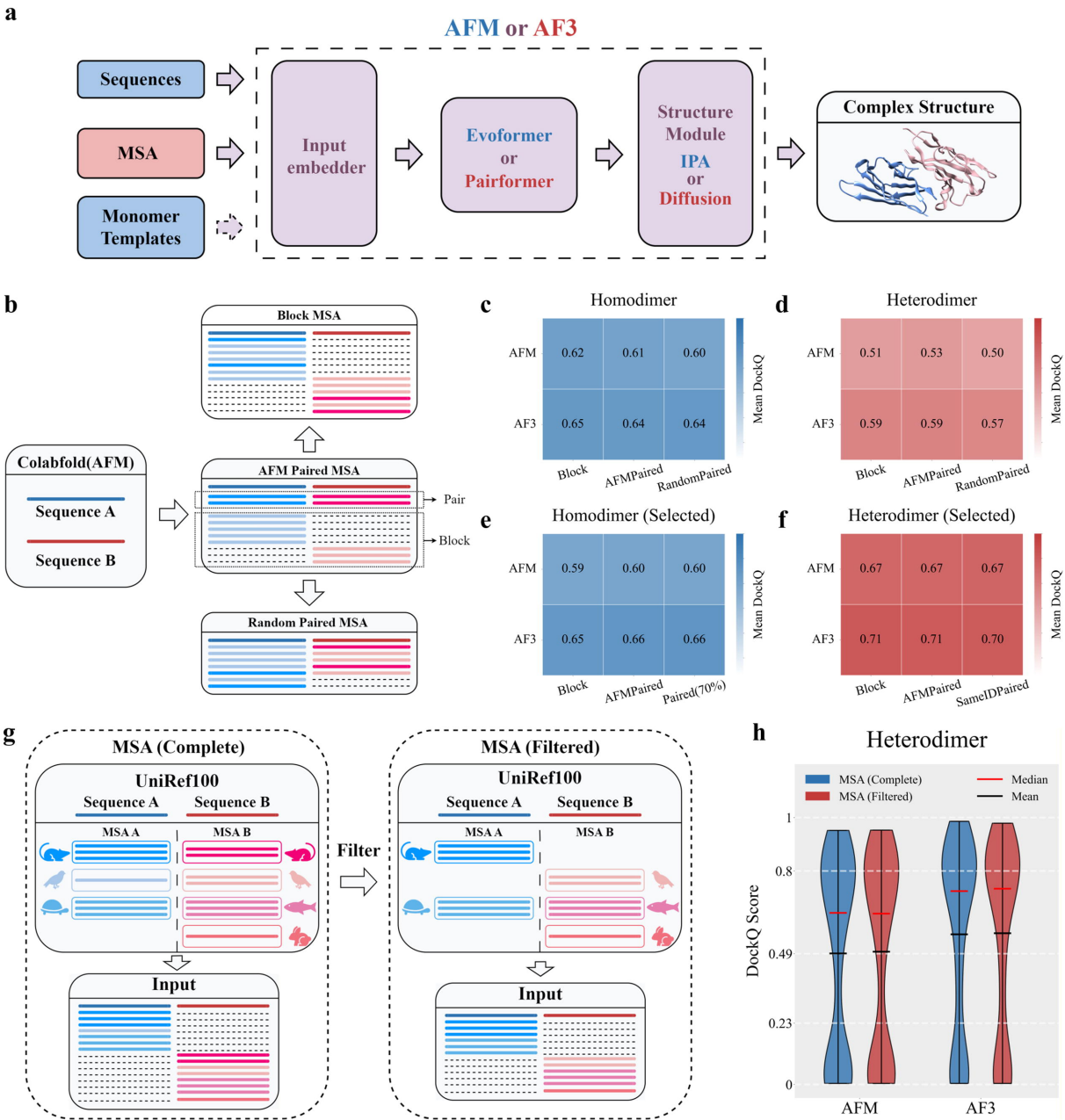


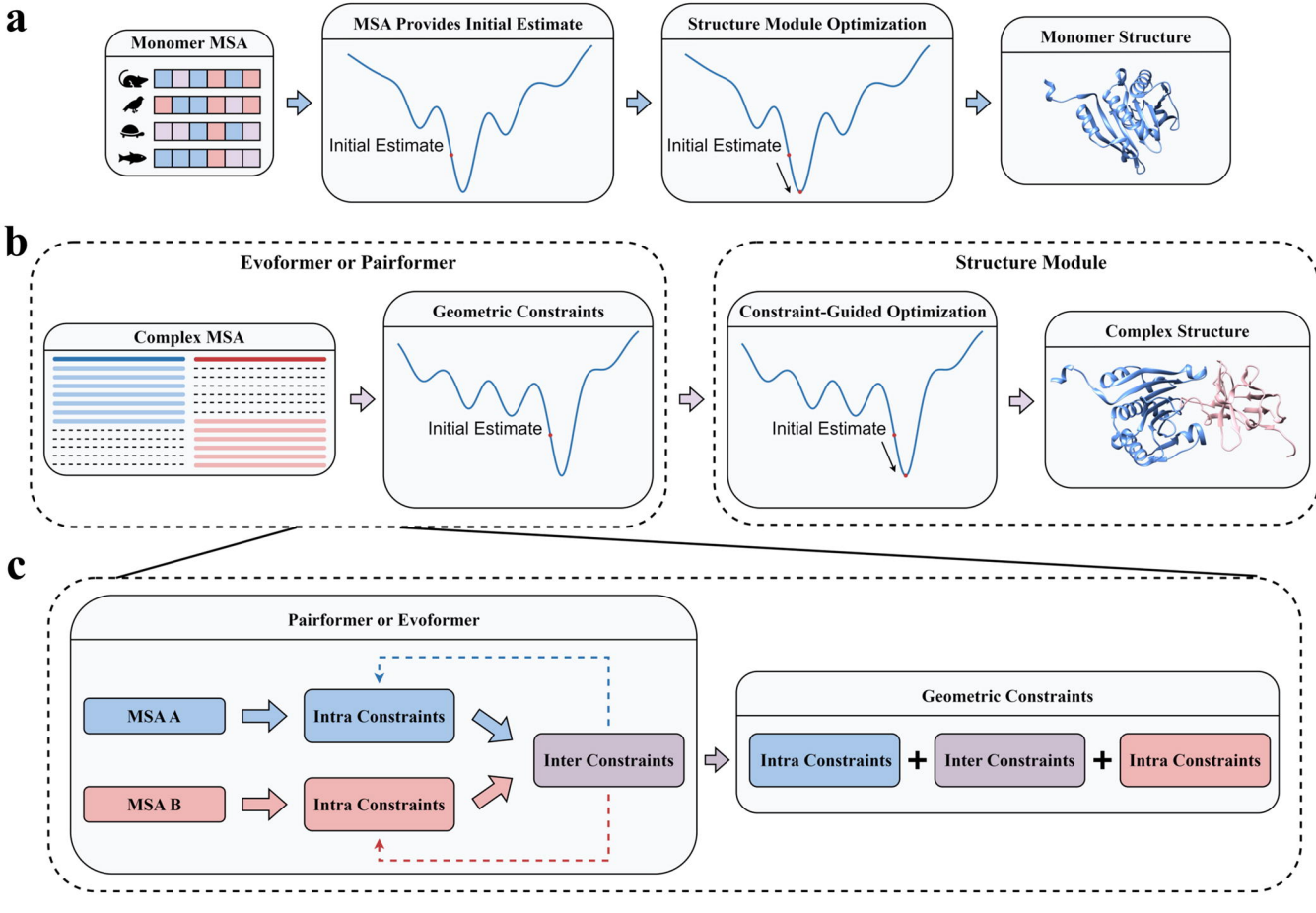
Figure 6. Mechanistic analysis of antigen–antibody prediction limitations.

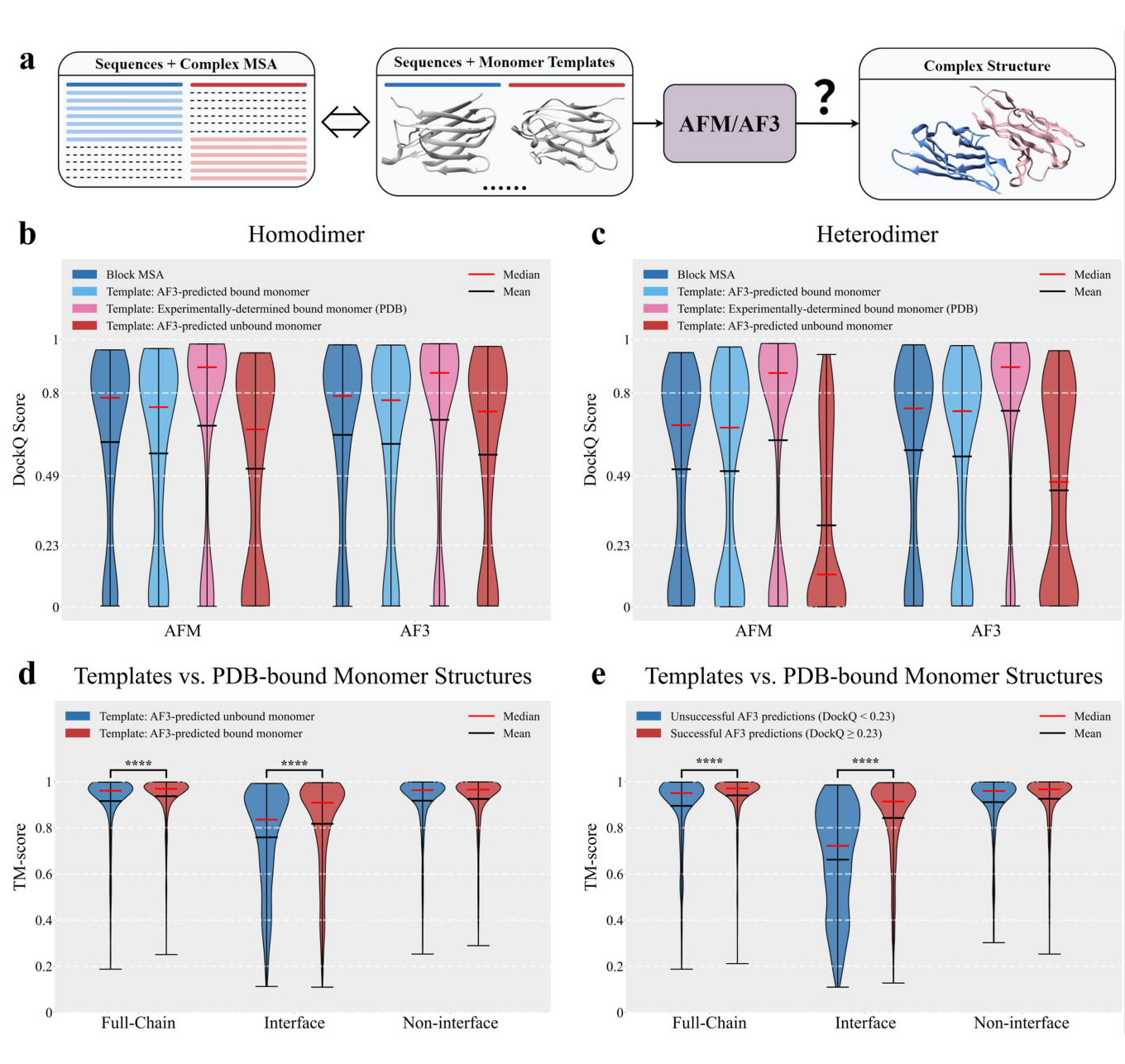
(a, b) Prediction performance of AFM and AF3 on the (a) antigen–antibody interface and (b) antibody heavy/light chain interface within the antigen–antibody dataset, evaluated under three input conditions: AFM MSA, Block MSA, and experimentally determined bound monomer templates (PDB).

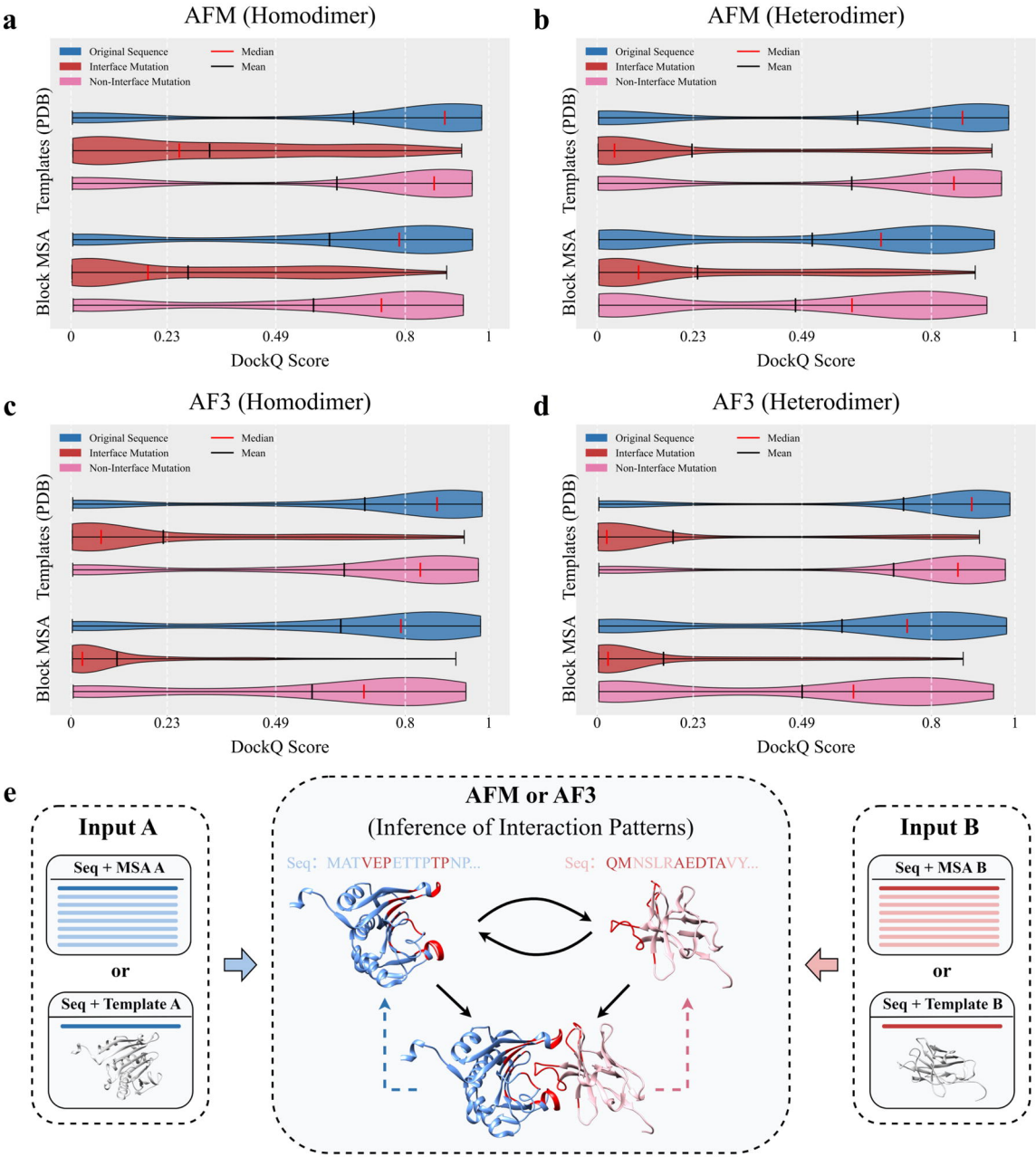
(c) Distribution of monomer TM-scores in AF3-predicted complex structures for heterodimers and antigen–antibody complexes at the full-chain, interface, and non-interface regions, using Block MSA as input, with interfaces defined from experimentally determined complex structures. At the full-chain and non-interface regions, antigen monomers show significantly lower accuracy than both antibody monomers and heterodimer monomers (****$p < 0.0001$, two-sided Mann–Whitney U test). At the interface region, both antigen and antibody monomers exhibit substantially lower accuracy than heterodimer monomers (****$p < 0.0001$).

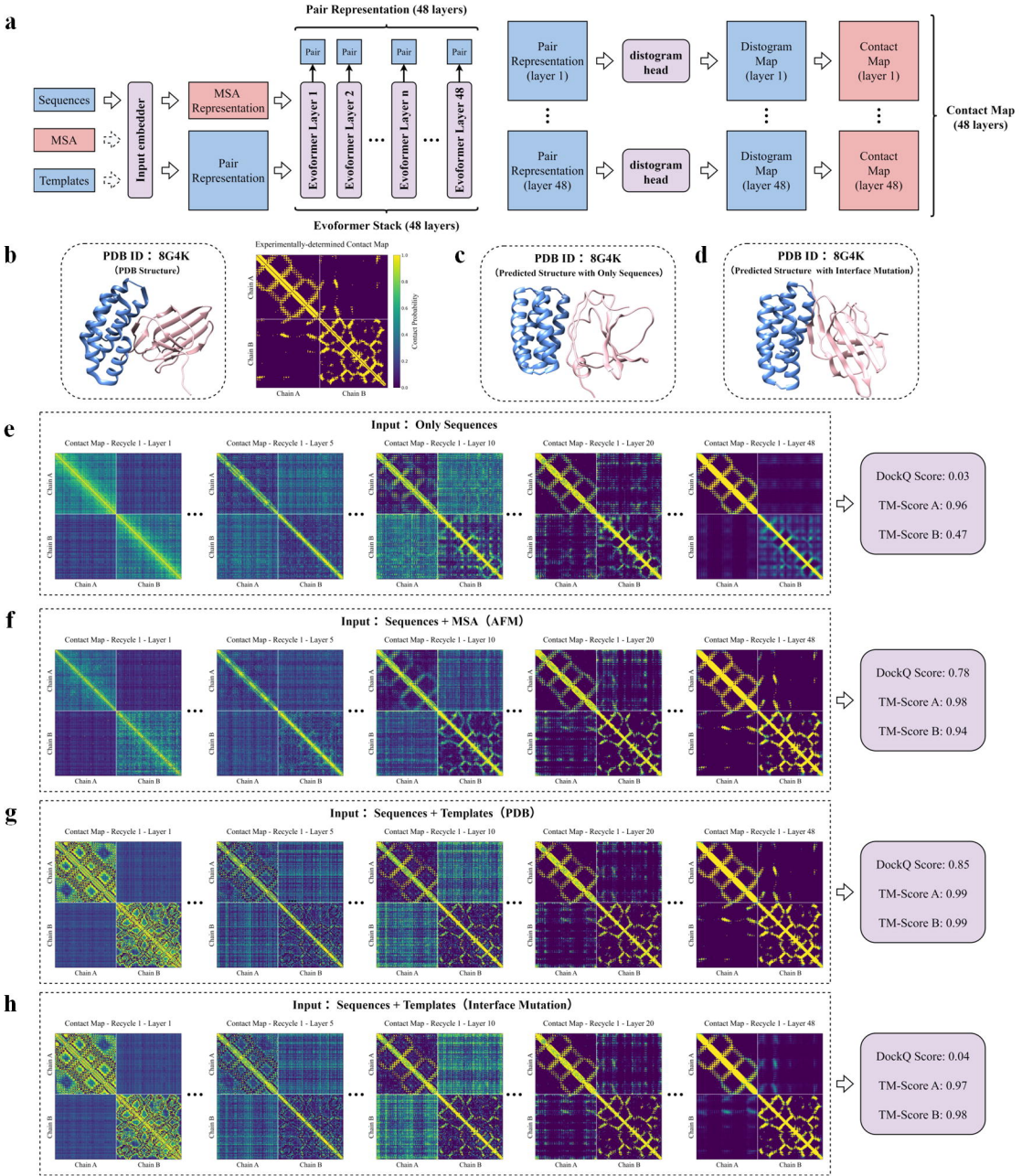
(d) Correlation between modeling accuracy (TM-score) of the three CDRs in antibody heavy and light chains and the prediction accuracy of the antigen–antibody interface (DockQ).
(e) Statistical analysis of residue-residue contact types at heterodimer and antigen–antibody interfaces.
(f) Experimental structure (left) and corresponding contact map (right) of 8BK2.
(g) Evolution of constraints and predicted model accuracy for 8BK2 under AFM MSA input with different recycling rounds (1 and 5).

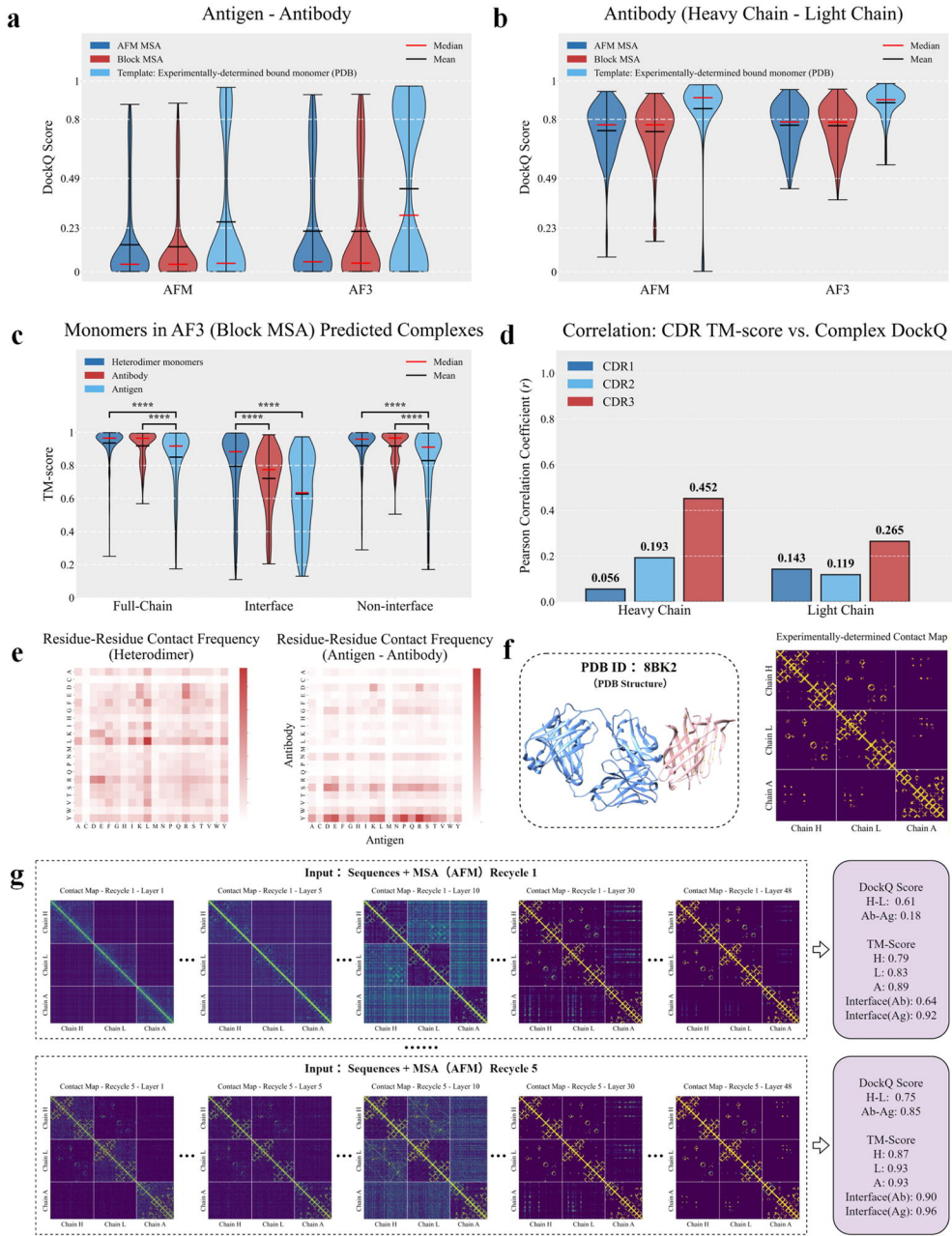

a
AFM or AF3
Sequences
MSA
Monomer Templates
Input embedder
Evoformer or Pairformer
Structure Module
IPA or Diffusion
Complex Structure
b
Colabfold(AFM)
Sequence A
Sequence B
Block MSA
AFM Paired MSA
Pair
Block
Random Paired MSA
c
Homodimer
AFM 0.62 0.61 0.60
AF3 0.65 0.64 0.64
Block AFMPaired RandomPaired
Mean DockQ
d
Heterodimer
AFM 0.51 0.53 0.50
AF3 0.59 0.59 0.57
Block AFMPaired RandomPaired
Mean DockQ
e
Homodimer (Selected)
AFM 0.59 0.60 0.60
AF3 0.65 0.66 0.66
Block AFMPaired Paired(70%)
Mean DockQ
f
Heterodimer (Selected)
AFM 0.67 0.67 0.67
AF3 0.71 0.71 0.70
Block AFMPaired SameIDPaired
Mean DockQ
g
MSA (Complete)
UniRef100
Sequence A
Sequence B
MSA A
MSA B
Input
Filter
MSA (Filtered)
UniRef100
Sequence A
Sequence B
MSA A
MSA B
Input
h
Heterodimer
MSA (Complete)
MSA (Filtered)
Median
Mean
DockQ Score
1
0.8
0.49
0.23
0
AFM
AF3


a
Monomer MSA
MSA Provides Initial Estimate
Initial Estimate
Structure Module Optimization
Initial Estimate
Monomer Structure
b
Evoformer or Pairformer
Complex MSA
Geometric Constraints
Initial Estimate
Structure Module
Constraint-Guided Optimization
Initial Estimate
Complex Structure
c
Pairformer or Evoformer
MSA A
Intra Constraints
MSA B
Intra Constraints
Inter Constraints
Geometric Constraints
Intra Constraints
+
Inter Constraints
+
Intra Constraints


a
Sequences + Complex MSA
Sequences + Monomer Templates
AFM/AF3
?
Complex Structure
b
Homodimer
Block MSA
Template: AF3-predicted bound monomer
Template: Experimentally-determined bound monomer (PDB)
Template: AF3-predicted unbound monomer
Median
Mean
DockQ Score
1
0.8
0.49
0.23
0
AFM
AF3
c
Heterodimer
Block MSA
Template: AF3-predicted bound monomer
Template: Experimentally-determined bound monomer (PDB)
Template: AF3-predicted unbound monomer
Median
Mean
DockQ Score
1
0.8
0.49
0.23
0
AFM
AF3
d
Templates vs. PDB-bound Monomer Structures
Template: AF3-predicted unbound monomer
Template: AF3-predicted bound monomer
Median
Mean
****
****
TM-score
1
0.8
0.6
0.4
0.2
0
Full-Chain
Interface
Non-interface
e
Templates vs. PDB-bound Monomer Structures
Unsuccessful AF3 predictions (DockQ < 0.23)
Successful AF3 predictions (DockQ ≥ 0.23)
Median
Mean
****
****
TM-score
1
0.8
0.6
0.4
0.2
0
Full-Chain
Interface
Non-interface


a
AFM (Homodimer)
b
AFM (Heterodimer)
c
AF3 (Homodimer)
d
AF3 (Heterodimer)
Original Sequence
Interface Mutation
Non-Interface Mutation
Median
Mean
Templates (PDB)
Block MSA
0
0.23
0.49
0.8
1
DockQ Score
e
Input A
Seq + MSA A
or
Seq + Template A
AFM or AF3
(Inference of Interaction Patterns)
Seq: MATVEPETTPTPNP...
Seq: QMNSLRAEDTAVY...
Input B
Seq + MSA B
or
Seq + Template B


a
Pair Representation (48 layers)
Sequences
MSA
Templates
Input embedder
MSA Representation
Pair Representation
Pair
Evoformer Layer 1
Evoformer Layer 2
Evoformer Layer n
Evoformer Layer 48
Evoformer Stack (48 layers)
Pair Representation (layer 1)
Pair Representation (layer 48)
distogram head
Distogram Map (layer 1)
Distogram Map (layer 48)
Contact Map (layer 1)
Contact Map (layer 48)
Contact Map (48 layers)
b
PDB ID：8G4K
(PDB Structure)
Experimentally-determined Contact Map
Chain A
Chain B
Contact Probability
c
PDB ID：8G4K
(Predicted Structure with Only Sequences)
d
PDB ID：8G4K
(Predicted Structure with Interface Mutation)
e
Input：Only Sequences
Contact Map - Recycle 1 - Layer 1
Contact Map - Recycle 1 - Layer 5
Contact Map - Recycle 1 - Layer 10
Contact Map - Recycle 1 - Layer 20
Contact Map - Recycle 1 - Layer 48
DockQ Score: 0.03
TM-Score A: 0.96
TM-Score B: 0.47
f
Input：Sequences + MSA（AFM）
DockQ Score: 0.78
TM-Score A: 0.98
TM-Score B: 0.94
g
Input：Sequences + Templates（PDB）
DockQ Score: 0.85
TM-Score A: 0.99
TM-Score B: 0.99
h
Input：Sequences + Templates（Interface Mutation）
DockQ Score: 0.04
TM-Score A: 0.97
TM-Score B: 0.98


a
Antigen - Antibody
AFM MSA
Block MSA
Template: Experimentally-determined bound monomer (PDB)
Median
Mean
DockQ Score
1
0.8
0.49
0.23
0
AFM
AF3
b
Antibody (Heavy Chain - Light Chain)
AFM MSA
Block MSA
Template: Experimentally-determined bound monomer (PDB)
Median
Mean
DockQ Score
1
0.8
0.49
0.23
0
AFM
AF3
c
Monomers in AF3 (Block MSA) Predicted Complexes
Heterodimer monomers
Antibody
Antigen
Median
Mean
TM-score
1
0.8
0.6
0.4
0.2
0
****
Full-Chain
Interface
Non-interface
d
Correlation: CDR TM-score vs. Complex DockQ
CDR1
CDR2
CDR3
Pearson Correlation Coefficient (r)
1.0
0.8
0.6
0.4
0.2
0.0
0.056
0.193
0.452
0.143
0.119
0.265
Heavy Chain
Light Chain
e
Residue-Residue Contact Frequency (Heterodimer)
Residue-Residue Contact Frequency (Antigen - Antibody)
Antibody
Antigen
f
PDB ID：8BK2
(PDB Structure)
Experimentally-determined Contact Map
Chain H
Chain L
Chain A
g
Input：Sequences + MSA（AFM）Recycle 1
Contact Map - Recycle 1 - Layer 1
Contact Map - Recycle 1 - Layer 5
Contact Map - Recycle 1 - Layer 10
Contact Map - Recycle 1 - Layer 30
Contact Map - Recycle 1 - Layer 48
DockQ Score
H-L: 0.61
Ab-Ag: 0.18
TM-Score
H: 0.79
L: 0.83
A: 0.89
Interface(Ab): 0.64
Interface(Ag): 0.92
Input：Sequences + MSA（AFM）Recycle 5
Contact Map - Recycle 5 - Layer 1
Contact Map - Recycle 5 - Layer 5
Contact Map - Recycle 5 - Layer 10
Contact Map - Recycle 5 - Layer 30
Contact Map - Recycle 5 - Layer 48
DockQ Score
H-L: 0.75
Ab-Ag: 0.85
TM-Score
H: 0.87
L: 0.93
A: 0.93
Interface(Ab): 0.90
Interface(Ag): 0.96