nature

https://doi.org/10.1038/s41586-024-07487-w

Accelerated Article Preview

Accurate structure prediction of biomolecular interactions with AlphaFold 3

Received: 19 December 2023

Accepted: 29 April 2024

Accelerated Article Preview

Cite this article as: Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* https://doi.org/10.1038/s41586-024-07487-w (2024)

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis & John M. Jumper

Accurate structure prediction of biomolecular interactions with AlphaFold 3

Josh Abramson[*1], Jonas Adler[*1], Jack Dunger[*1], Richard Evans[*1], Tim Green[*1], Alexander Pritzel[*1], Olaf Ronneberger[*1], Lindsay Willmore[*1], Andrew J Ballard[1], Joshua Bambrick[2], Sebastian W Bodenstein[1], David A Evans[1], Chia-Chun Hung[2], Michael O'Neill[1], David Reiman[1], Kathryn Tunyasuvunakool[1], Zachary Wu[1], Akvilė Žemgulytė[1], Eirini Arvaniti[3], Charles Beattie[3], Ottavia Bertolli[3], Alex Bridgland[3], Alexey Cherepanov[4], Miles Congreve[4], Alexander I Cowen-Rivers[3], Andrew Cowie[3], Michael Figurnov[3], Fabian B Fuchs[3], Hannah Gladman[3], Rishub Jain[3], Yousuf A Khan[3], Caroline M R Low[4], Kuba Perlin[3], Anna Potapenko[3], Pascal Savy[4], Sukhdeep Singh[3], Adrian Stecula[4], Ashok Thillaisundaram[3], Catherine Tong[4], Sergei Yakneen[4], Ellen D Zhong[3], Michal Zielinski[3], Augustin Žídek[3], Victor Bapst[†1], Pushmeet Kohli[†1], Max Jaderberg[†2], Demis Hassabis[†1,2], John M Jumper[†1]

[*]Contributed equally
[1]Core Contributor, Google DeepMind, London, UK
[2]Core Contributor, Isomorphic Labs, London, UK
[3]Google DeepMind, London, UK
[4]Isomorphic Labs, London, UK
[†] Jointly supervised

Corresponding author emails:
J. J. - *jumper@google.com*; D.H. - *dhcontact@google.com;* M.J. - *jaderberg@isomorphiclabs.com*

The introduction of AlphaFold 2[1] has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein modelling and design[2–6]. In this paper, we describe our AlphaFold 3 model with a substantially updated diffusion-based architecture, which is capable of joint structure prediction of complexes including proteins, nucleic acids, small molecules, ions, and modified residues. The new AlphaFold model demonstrates significantly improved accuracy over many previous specialised tools: far greater accuracy on protein-ligand interactions than state of the art docking tools, much higher accuracy on protein-nucleic acid interactions than nucleic-acid-specific predictors, and significantly higher antibody-antigen prediction accuracy than AlphaFold-Multimer v2.3[7,8]. Together these results show that high accuracy modelling across biomolecular space is possible within a single unified deep learning framework.

Main Text

Introduction

Accurate models of biological complexes are critical to our understanding of cellular functions and for the rational design of therapeutics[2–4,9]. Enormous progress has been achieved in protein structure prediction with the development of AlphaFold[1], and the field has grown tremendously with a number of later methods that build on the ideas and techniques of AlphaFold 2[10–12]. Almost immediately after AlphaFold became available, it was shown that simple input modifications would enable surprisingly accurate protein interaction predictions[13–15] and that training AlphaFold 2 specifically for protein interaction prediction yielded a highly accurate system[7].

These successes lead to the question of whether it is possible to accurately predict the structure of complexes containing a much wider range of biomolecules, including ligands, ions, nucleic acids, and modified residues, within a deep learning framework. A wide range of predictors for various specific interaction types have been developed[16–28], as well as one generalist method developed concurrently with the present work[29], but the accuracy of such deep learning attempts has been mixed and often below that of physics-inspired methods[30,31]. Almost all these methods are also highly specialised to particular interaction types and cannot predict the structure of general biomolecular complexes containing many types of entities.

Here, we present AlphaFold 3 (AF3), a model that is capable of high accuracy prediction of complexes containing nearly all molecular types present in the Protein Data Bank[32] (PDB) (**Fig. 1a,b**). In all but one category it achieves a significantly higher performance than strong methods that specialise in just the given task (**Fig. 1c, Extended Data Table 1**) including higher accuracy at protein structure and the structure of protein-protein interactions.

This is achieved by a substantial evolution of the AlphaFold 2 architecture and training procedure (**Fig. 1d**) both to accommodate more general chemical structures and to improve the data efficiency of learning. The system reduces the amount of multiple sequence alignment (MSA) processing by replacing the AlphaFold 2 Evoformer with the simpler Pairformer Module (**Fig. 2a)**. Furthermore it directly predicts the raw atom coordinates with a Diffusion Module, replacing the AlphaFold 2 Structure Module that operated on amino-acid-specific frames and side chain torsion angles (**Fig. 2b**). The multiscale nature of the diffusion process (low noise levels induce the network to improve local structure) also allow us to eliminate stereochemical losses and most special handling of bonding patterns in the network, easily accommodating arbitrary chemical components.

Network architecture and training

The overall structure of AF3 (**Fig. 1d, Supplementary Methods 3**) echoes that of AlphaFold 2 with a large trunk evolving a pairwise representation of the chemical complex followed by a Structure Module that uses the pairwise representation to generate explicit atomic positions, but there are large differences in each major component. These modifications were driven both by the need to accommodate a wide range of chemical entities without excessive special-casing and by observations of AlphaFold 2 performance with different modifications. Within the trunk, MSA processing is substantially de-emphasized with a much smaller and simpler MSA embedding block (**Supplementary Methods 3.3**). Compared to the original Evoformer from AlphaFold 2 the number of blocks are reduced to four, the processing of the MSA representation uses an inexpensive pair-weighted averaging, and only the pair representation is used for later processing steps. The "Pairformer" (**Fig. 2a, Supplementary Methods 3.6)** replaces the "Evoformer" of AlphaFold 2 as the dominant processing block. It operates only on the pair representation and the single representation; the MSA representation is not retained and all information passes via the pair representation. The pair processing and the number of blocks (48) is largely unchanged from AlphaFold 2. The resulting pair and single representation together with the input representation are passed to the new Diffusion Module (**Fig. 2b)** that replaces the Structure Module of AlphaFold 2.

The Diffusion Module (**Fig. 2b, Supplementary Methods 3.7)** operates directly on raw atom coordinates, and on a coarse abstract token representation, without rotational frames or any equivariant processing. We had observed in AlphaFold 2 that removing most of the complexity of the Structure Module had only a modest effect on prediction accuracy, and maintaining the backbone frame and side chain torsion representation add quite a bit of complexity for general molecular graphs. Similarly AlphaFold 2 required carefully tuned stereochemical violation penalties during training to enforce chemical plausibility of the resulting structures. We use a relatively standard diffusion approach[34] in which the diffusion model is trained to receive "noised" atomic coordinates then predict the true coordinates. This task requires the network to learn protein structure at a variety of length scales, where the denoising task at small noise emphasises understanding very local stereochemistry and the denoising task at high noise emphasises large-scale structure of the system. At inference time, random noise is sampled and then recurrently denoised to produce a final structure. Importantly, this is a generative training procedure which produces a distribution of answers. This means that, for each answer, the local structure will be sharply defined (e.g. side chain bond geometry) even when the network is uncertain about the positions. For this reason, we are able to avoid both torsion-based parametrizations of the residues and violation losses on the structure, while handling the full complexity of general ligands. Similarly to some recent work[35], we find that no invariance or equivariance with respect to global rotations and translation of the molecule are required in the architecture and so we omit them to simplify the machine learning architecture.

The use of a generative diffusion approach comes with some technical challenges that we needed to address. The biggest issue is that generative models are prone to hallucination[36] where the

model may invent plausible-looking structure even in unstructured regions. To counteract this effect, we use a novel cross-distillation method where we enrich the training data with AlphaFold-Multimer v2.3[7,8] predicted structures. In these structures, unstructured regions are typically represented by long extended loops instead of compact structures and training on them "teaches" AlphaFold 3 to mimic this behaviour. This cross-distillation greatly reduced the hallucination behaviour of AF3 (**Extended Data Fig. 1** for disorder prediction results on the CAID 2[37] benchmark set).

We also developed confidence measures that predict the atom-level and pairwise errors in our final structures. In AlphaFold 2, this was done directly by regressing the error in the output of the Structure Module during training. This procedure is not applicable to diffusion training however, since only a single step of the diffusion is trained instead of a full structure generation (**Fig. 2c**). To remedy this, we developed a diffusion "rollout" procedure for the full structure prediction generation during training (using a larger step size than normal; see **Fig. 2c** "mini-rollout"). This predicted structure is then used to permute the symmetric ground truth chains and ligands, and to compute the performance metrics to train the confidence head. The confidence head uses the pairwise representation to predict the LDDT (pLDDT) and a predicted aligned error (PAE) matrix as in AlphaFold 2, as well as a distance error matrix (PDE) which is the error in the distance matrix of the predicted structure as compared to the true structure (see **Supplementary Methods 4.3** for details).

Fig. 2d shows that during initial training the model learns quickly to predict the local structures (all intra chain metrics go up quickly and reach 97% of the maximum performance within the first 20k training steps) while the model needs considerably longer to learn the global constellation (the interface metrics go up slowly and protein-protein interface LDDT passes the 97% bar only after 60k steps). During AF3 development we observed that some model capabilities topped out relatively early and started to decline (most likely due to overfitting to the limited number of training samples for this capability) while other capabilities were still undertrained. We addressed this by increasing / decreasing the sampling probability for the corresponding training sets (**Supplementary Methods 2.5.1**) and by an early stopping using a weighted average of all above metrics and some additional metrics to select the best model checkpoint (**Supplementary Table 7**). The fine tuning stages with the larger crop sizes improve the model on all metrics with an especially high uplift on protein-protein interfaces (**Extended Data Fig. 2**).

Accuracy across complex types

AF3 can predict structures from input polymer sequences, residue modifications, and ligand SMILES. In **Fig. 3** we show a selection of examples highlighting the ability of the model to generalise to a number of biologically important and therapeutically relevant modalities. In selecting these examples, we considered novelty in terms of the similarity of individual chains and interfaces to the training set (additional information in **Supplementary Methods 8.1)**.

We evaluate performance of the system on recent interface-specific benchmarks for each complex type (**Fig. 1c, Extended Data Table 1).** Performance on protein-ligand interfaces was evaluated on the PoseBusters benchmark set, composed of 428 protein-ligand structures released to the PDB in 2021 or later. Since our standard training cutoff date is in 2021, we trained a separate AF3 model with an earlier training set cutoff (see **Methods** for details). Accuracy on the PoseBusters set is reported as the percentage of protein-ligand pairs with pocket-aligned ligand RMSD of less than 2 Å. The baseline models come in two categories: those that use only protein sequence and ligand SMILES as input and those that additionally leak information from the solved protein-ligand test structure. Traditional docking methods use the latter privileged information, even though that information would not be available in real world use cases. Even so, AlphaFold 3 greatly outperforms classical docking tools like Vina[38,39] even while not using any structural inputs (Fisher exact $p=2.27 * 10^{-13}$) and greatly outperforms all other true blind docking like RoseTTAFold All-Atom ($p=4.45 * 10^{-25}$). **Extended Data Fig. 3** shows three examples where AlphaFold 3 achieves accurate predictions but docking tools Vina and Gold do not[38]. PoseBusters analysis was done using a 2019-09-30 training cutoff for AlphaFold 3 to ensure the model was not trained on any PoseBusters structures. To compare to RoseTTAFold All-Atom results, we used PoseBusters Version 1. Version 2 (crystal contacts removed from the benchmark set) results including quality metrics are shown in **Extended Data Fig. 4b-f** and in **Extended Data Table 1.** We use multiple seeds to ensure correct chirality and avoid slight protein-ligand clashing (as opposed to a method like diffusion guidance to enforce) but are typically able to produce high quality stereochemistry. Separately, we also train a version of AlphaFold 3 that receives the "pocket information" as used in some recent deep learning work[24,26] (**Extended Data Fig. 4a** for results).

AF3 predicts protein-nucleic complexes and RNA structures with higher accuracy than RoseTTAFold2NA[40] (**Fig. 1c** second plot**).** As RoseTTAFold2NA is only validated on structures below 1000 residues, we use only structures below 1000 residues from our Recent PDB evaluation set for this comparison (see **Methods** for details). AlphaFold 3 is able to predict protein-nucleic structures with thousands of residues, an example of which is shown in **Fig. 3a**. Note that we do not compare directly to RoseTTAFold All-Atom, but benchmarks indicate that RoseTTAFold All-Atom is comparable to slightly less accurate than RoseTTAFold2NA for nucleic acid predictions[29].

We also evaluated AF3 performance on the 10 publicly available CASP15 RNA targets: We achieve a higher average performance than RoseTTAFold2NA and AIchemy_RNA[27] (the best AI-based submission in CASP15[18,31]) on the respective common subsets of our and their predictions (see **Extended Data Fig. 5a** for detailed results). We do not reach the performance of the best human-expert-aided CASP15 submission AIchemy_RNA2[41] (**Fig. 1c**, centre left). Due to limited dataset sizes, we do not report significance test statistics here. Further analysis of the accuracy of predicting nucleic acids alone (without proteins) is shown in **Extended Data Fig. 5b**.

Covalent modifications (bonded ligands, glycosylation, and modified protein residues and nucleic acid bases) are also accurately predicted by AF3 (**Fig. 1c**, centre right). Modifications include those to any polymer residue (protein, RNA or DNA). We report accuracy as the percentage of successful predictions (pocket RMSD < 2 Å). We apply quality-filters to the bonded ligands and glycosylation dataset (as does PoseBusters): We only include ligands with high-quality experimental data (ranking_model_fit > 0.5 according to the RCSB structure validation report, that is, X-ray structures with a model quality above the median). As with the PoseBusters set, the bonded ligands and glycosylation datasets are not filtered by homology to the training data set. Filtering based on the bound polymer chain homology (using polymer template similarity < 40) yielded only 5 clusters for bonded ligands and 7 clusters for glycosylation. We exclude multi-residue glycans here, because the RCSB validation report does not provide a ranking_model_fit value for them. The percentage of successful predictions (pocket RMSD < 2 Å) for multi-residue glycans on all-quality experimental data is 42.1% (N=131 clusters) which is slightly lower than the success rate for single-residue glycans on all-quality experimental data of 46.1% (N=167). The modified residues dataset is filtered similarly to our other polymer test sets: it contains only modified residues in polymer chains with low homology to the training set (see **Methods** for details). See **Extended Data Table 1** for detailed results; **Extended Data Fig. 6** for examples of predicted protein, DNA, and RNA structures with covalent modifications including analysis of the impact phosphorylation has on predictions.

While expanding in modelling capabilities, AF3 has also improved in protein complex accuracy relative to AlphaFold-Multimer v2.3[7,8] (AF-M 2.3). Generally, protein-protein prediction success (DockQ[42] > 0.23) has increased (paired Wilcoxon signed-rank test, $p=1.8 * 10^{-18}$), with antibody-protein interaction prediction in particular showing a marked improvement (**Fig. 1c** right, paired Wilcoxon signed-rank test, $p=6.5 * 10^{-5}$, predictions top ranked from 1000 rather than the typical 5 seeds, see **Fig. 5a** for details). Protein monomer LDDT improvement is also significant (paired Wilcoxon signed-rank test, $p=1.7 * 10^{-34}$). AF3 has a very similar dependence on MSA depth to AF-M 2.3; proteins with shallow MSAs are predicted with lower accuracy (see **Extended Data Fig. 7a** for a comparison of the dependence of single chain LDDT on MSA depth).

Predicted confidences track accuracy

As with AlphaFold 2, AlphaFold 3 confidence measures are well-calibrated with accuracy. Our confidence analysis is performed on the recent PDB evaluation set, with no homology filtering and including peptides. The ligands category is filtered to high quality experimental structures as described above, and considers standard non-bonded ligands only. See **Extended Data Fig. 8** for a similar assessment on bonded ligand and other interfaces. All statistics are cluster-weighted (see **Methods** for details) and consider the top-ranked prediction only (see **Supplementary Methods 5.9.3** for ranking details).

In **Fig. 4a** top row we plot chain pair ipTM (interface predicted TM score[43]; see **Supplementary Methods 5.9.1**) against interface accuracy measures: protein-protein DockQ, protein-nucleic iLDDT, and protein-ligand success, with success defined as percent of examples under thresholded pocked-aligned RMSD values. In **Fig. 4a** bottom row we plot average pLDDT per protein, nucleotide or ligand entity against our bespoke LDDT_to_polymer metric (for metrics details, see **Methods**), which is closely related to the training target of the pLDDT predictor.

In **Fig. 4b-e** we highlight a single example prediction of 7T82, where per-atom pLDDT colouring identifies unconfident chain tails, somewhat confident interfaces, and otherwise confident secondary structure. In **Fig. 4c** the same prediction is coloured by chain, along with DockQ interface scores in **Fig. 4d** and per-chain colouring displayed on the axes for reference. We see from **Fig. 4e** that PAE confidence is high for pink-grey and blue-orange residue pairs where DockQ > 0.7, and least confident about pink-orange and pink-blue residue pairs which have DockQ ≈ 0. See **Extended Data Fig. 5c-d** for a similar PAE analysis on an example with protein and nucleic acid chains.

Model limitations

We note model limitations of AlphaFold 3 with respect to stereochemistry, hallucinations, dynamics, and accuracy for certain targets.

On stereochemistry, we note two main classes of violations. The first is that the model outputs do not always respect chirality (**Fig. 5b**), despite the model receiving reference structures with correct chirality as input features. To address this in the PoseBusters benchmark, we included a penalty for chirality violation in our ranking formula for model predictions. Despite this, we still observe a chirality violation rate of 4.4% in the benchmark. The second class of stereochemical violations is a tendency of the model to occasionally produce overlapping ("clashing") atoms in the predictions. This sometimes manifests as extreme violations in homomers where entire chains have been observed to overlap (**Fig. 5e**). Penalising clashes during ranking (see **Supplementary Methods 5.9.3**) reduces the occurrence of this failure mode but does not eliminate them. Almost all remaining clashes occur for protein-nucleic complexes with both greater than 100 nucleotides and greater than 2,000 residues in total.

We note that the switch from the non-generative AlphaFold 2 model to the diffusion-based AlphaFold 3 model introduces the challenge of spurious structural order (hallucinations) in disordered regions (**Fig. 5d**, **Extended Data Fig. 1**). While hallucinated regions are typically marked as very low confidence, they can lack the distinctive ribbon-like appearance that AlphaFold 2 produces in disordered regions. To encourage ribbon-like predictions in AF3, we use distillation training from AlphaFold 2 predictions, and we add a ranking term to encourage results with more solvent accessible surface area[37].

A key limitation of protein structure prediction models is that they typically predict static structures as seen in the PDB, not the dynamical behaviour of biomolecular systems in solution. This limitation persists for AlphaFold 3, where multiple random seeds for either the diffusion head or the overall network do not produce an approximation of the solution ensemble.

In some cases, the modelled conformational state may not be correct or comprehensive given the specified ligands and other inputs. As an example, E3 ubiquitin ligases natively adopt an open conformation in an apo state and have only been observed in a closed state when bound to ligands, but AF3 exclusively predicts the closed state for both holo and apo systems[44] (**Fig. 5c**). Many methods have been developed, particularly around MSA resampling, which assist in generating diversity from previous AlphaFold models[45–47] and may also assist in multi-state prediction with AF3.

Despite the large advance in modelling accuracy in AlphaFold 3, there are still many targets for which accurate modelling can be challenging. To obtain the highest accuracy, it may be necessary to generate a large number of predictions and rank them, which incurs an extra computational cost. A class of targets where we observe this effect strongly is antibody-antigen complexes similar to other recent work[48]. **Fig. 5a** shows that for AlphaFold 3, top-ranked predictions keep improving with more model seeds, even at as many as 1000 (Wilcoxon signed rank test between 5 and 1000 seeds, $p=2.0 * 10^{-5}$ for % correct and $p=0.009$ for % very high accuracy; ranking by protein-protein interface ipTM). This large improvement with many seeds isn't observed in general for other classes of molecules (see **Extended Data Fig. 7b).** Using only one diffusion sample per model seed for the AF3 predictions rather than five (not illustrated) does not change results significantly, indicating that running more model seeds is necessary for antibody score improvements, rather than just more diffusion samples.

Discussion

The core challenge of molecular biology is to understand and ultimately regulate the complex atomic interactions of biological systems. The AlphaFold 3 model takes a large step in this direction, demonstrating that it is possible to accurately predict the structure of a wide range of biomolecular systems in a unified framework. While there are still substantial challenges to achieve highly accurate predictions across all interaction types, we demonstrate that it is possible to build a deep learning system that shows strong coverage and generalisation for all these interactions. We also demonstrate that the lack of cross-entity evolutionary information is not a substantial blocker to progress in predicting these interactions, and moreover substantial improvement in antibody results suggests AlphaFold-derived methods are able to model the chemistry and physics of classes of molecular interactions without dependence on MSAs. Finally, the large improvement in protein-ligand structure prediction shows that it is possible to handle the wide diversity of chemical space within a general deep learning framework and without resorting to an artificial separation between protein structure prediction and ligand docking.

The development of bottom-up modelling of cellular components is a key step in unravelling the complexity of molecular regulation within the cell, and the performance of AlphaFold 3 shows that developing the right deep learning frameworks can massively reduce the amount of data required to obtain biologically relevant performance on these tasks and amplify the impact of the data already collected. We expect that structural modelling will continue to improve not only due to deep learning advances but also because continuing methodological advances in experimental structure determination, such as the dramatic improvements in cryo electron microscopy and tomography, will provide a wealth of new training data to further the improve the generalisation capability of such models. The parallel developments of experimental and computational methods promise to propel us further into an era of structurally informed biological understanding and therapeutic development.

Figure Captions

Fig. 1 | AlphaFold 3 accurately predicts structures across biomolecular complexes. a, Example structure predicted with AF3: bacterial CRP/FNR family transcriptional regulator protein bound to DNA and cGMP (PDB ID 7PZB, full complex LDDT[33]: 82.8, GDT: 90.1). **b,** Example structure predicted with AF3: human coronavirus OC43 spike protein, 4665 residues, heavily glycosylated and bound by neutralising antibodies (PDB ID 7PNM, full complex LDDT: 83.0, GDT: 83.1). **c,** Performance on PoseBusters (V1, August 2023 release), our Recent PDB evaluation set, and CASP15 RNA. Metrics are % of pocket-aligned ligand RMSD < 2 Å for ligands and covalent modifications, interface LDDT for protein-nucleic acid complexes, LDDT for nucleic acid and protein monomers, and % DockQ > 0.23 for protein-protein and protein-antibody interfaces. All scores are reported from the top confidence-ranked sample out of 5 model seeds (each with 5 diffusion samples), except for protein-antibody scores which were ranked across 1000 model seeds for both models (each AF3 seed with 5 diffusion samples). See **Methods** for sampling and ranking details. For ligands, N indicates number of targets; for nucleic acids, N indicates number of structures; for modifications, N indicates clusters, and for proteins N indicates clusters. Bar heights indicate means; error bars indicate exact binomial distribution 95% confidence intervals for PoseBusters and via 10,000 bootstrap resamples for all others. Significance levels calculated via two-sided Fisher's Exact Test for PoseBusters and via two-sided Wilcoxon signed rank test for all others. *** for $p < 0.001$, ** for $p<0.01$. P-values (left to right): $2.27*10^{-13}$, $2.57*10^{-3}$, $2.78*10^{-3}$, $7.28*10^{-12}$, $1.81*10^{-18}$, $6.54*10^{-5}$, and $1.74*10^{-34}$. **d,** AF3 architecture for inference. Rectangles represent processing modules, arrows show the data flow. yellow: input data, blue: abstract network activations, green: output data. Coloured balls represent physical atom coordinates.

Fig. 2 | Architectural and training details. a, Pairformer Module. Input and output: pair representation with dimension (n, n, c) and single representation with dimension (n, c). n: number of tokens (polymer residues and atoms), c: number of channels (128 for the pair representation, 384 for the single representation). Each of the 48 blocks has an independent set of trainable parameters. **b, Diffusion Module.** Input: coarse arrays depict per-token representations (green: inputs, blue: pair, red: single). Fine arrays depict per-atom representations. Coloured balls represent physical atom coordinates. **c,** training setup (distogram head omitted) starting from the end of the network trunk. Coloured arrays: activations from the network trunk (green: inputs, blue: pair, red: single). Blue arrows: abstract activation arrays; yellow arrows: ground truth data; green arrows: predicted data. Stop sign: stop gradient operation. Both depicted Diffusion Modules share weights. **d,** Training curves for initial training and fine tuning stages, showing LDDT on our evaluation set as a function of optimizer steps. The scatter plot shows the raw data points and the lines show the smoothed performance using a median filter with a kernel width of 9 data points. The crosses mark the point where the smoothed performance reaches 97% of its initial training maximum.

Fig. 3 | Examples of predicted complexes. : Selected structure predictions from AF3. Predicted protein chains are shown in blue (predicted antibody in green), predicted ligands and glycans in orange, predicted RNA in purple, and ground truth in grey**. a,** Human 40S small ribosomal subunit (7663 residues) including 18S ribosomal RNA and Met-$tRNA_i^{Met}$ (opaque purple) in complex with translation initiation factors eIF1A and eIF5B (opaque blue; PDB ID 7TQL, full complex LDDT: 87.7, GDT: 86.9). **b,**Glycosylated globular portion of an EXTL3 homodimer (PDB ID 7AU2, mean pocked-aligned RMSD: 1.10 Å). **c,** Mesothelin C-terminal peptide bound to the monoclonal antibody 15B6 (PDB ID 7U8C, DockQ: 0.85). **d**, LGK974, a clinical stage inhibitor, bound to PORCN in complex with the WNT3A peptide (PDB ID 7URD, ligand RMSD 1.00 Å). **e**, (5S,6S)-O7-sulfo DADH bound to the AziU3/U2 complex with a novel fold (PDB ID 7WUX, ligand RMSD 1.92 Å). **f**, Analog of NIH-12848 bound to an allosteric site of PI5P4Kγ (PDB ID 7QIE, ligand RMSD 0.37 Å).

Fig. 4 | AlphaFold 3 confidences track accuracy. **a** (top row), Accuracy of protein-containing interfaces as a function of chain pair ipTM. **a** (bottom row), LDDT_to_polymer accuracy evaluated for various chain types as a function of chain-averaged pLDDT. Box, centerline, and whiskers boundaries are at (25%, 75%) intervals, median, and (5%, 95%) intervals. N values report the number of clusters in each band. **b**, Predicted structure of PDB ID 7T82 coloured by pLDDT (orange: 0-50, yellow: 50-70, cyan 70-90, and blue 90-100). **c**, same prediction coloured by chain. **d**, DockQ scores for protein-protein interfaces. **e**, Predicted Aligned Error (PAE) matrix of same prediction (darker is more confident), with chain colouring of panel c on side-bars. Dashed black lines indicate chain boundaries.

Fig. 5 | Model limitations. a, Antibody prediction quality increases with the number of model seeds. Quality of top-ranked, low homology, antibody-antigen interface predictions as a function of number of seeds. Each datapoint shows the mean over 1,000 random samples (with replacement) of seeds to rank over, out of 1200 seeds. Confidence intervals are 95% bootstraps over 10,000 resamples of cluster scores at each datapoint. Samples per interface ranked by protein-protein ipTM. Significance tests are by a two-sided Wilcoxon signed rank test. N = 65 clusters. *** for $p < 0.001$. P-values: $2.0 * 10^{-5}$ for % correct and p=0.009 for % very high accuracy. **b**, Prediction (coloured) and ground truth (grey) structures of *Thermotoga maritima* alpha-glucuronidase and beta-D-glucuronic acid, a target from the PoseBusters set (PDB ID 7CTM). AF3 predicts alpha-D-glucuronic acid, differing chiral centre indicated by an asterisk. The prediction shown is top-ranked by ligand-protein ipTM and with a chirality and clash penalty. **c**, Conformation coverage is limited. Ground truth structures (grey) of cereblon in open (apo PDB ID 8CVP, left) and closed (holo mezigdomide-bound, PDB ID 8D7U, right) conformations. Predictions (blue) of both apo (with 10 overlaid samples) and holo structures are in the closed conformation. Dashed line indicates distance between the N-terminal Lon protease-like and C-terminal thalidomide-binding domain. **d**, A nuclear pore complex with 1,854 unresolved residues (PDB ID 7F60). Ground truth (left) and predictions from AF-M 2.3 (middle) and AF3 (right). **e**, Prediction of a trinucleosome with overlapping DNA (pink) and protein (blue) chains (PDB ID 7PEU); highlighted are overlapping protein chains B and J and self-overlapping DNA chain AA. Unless otherwise stated, predictions are top-ranked by our global complex ranking metric with chiral mismatch and steric clash penalties (see **Supplementary Methods 5.9.1**).

Citations

1. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

2. Kreitz, J. *et al.* Programmable protein delivery with a bacterial contractile injection system. *Nature* **616**, 357–364 (2023).

3. Lim, Y. *et al.* In silico protein interaction screening uncovers DONSON's role in replication initiation. *Science* **381**, eadi3448 (2023).

4. Mosalaganti, S. *et al.* AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* **376**, eabm9506 (2022).

5. Anand, N. & Achim, T. Protein Structure and Sequence Generation with Equivariant Denoising Diffusion Probabilistic Models. arXiv preprint arXiv:2205.15019 (2022) doi:10.48550/arXiv.2205.15019.

6. Yang, Z., Zeng, X., Zhao, Y. & Chen, R. AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduction and Targeted Therapy* **8**, 1–14 (2023).

7. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 (2022) doi:10.1101/2021.10.04.463034.

8. alphafold/docs/technical_note_v2.3.0.md at main · google-deepmind/alphafold. *GitHub* https://github.com/google-deepmind/alphafold/blob/main/docs/technical_note_v2.3.0.md.

9. Structure-based drug design with geometric deep learning. *Curr. Opin. Struct. Biol.* **79**, 102548 (2023).

10. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).

11. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* (2021) doi:10.1126/science.abj8754.

12. Wu, R. *et al.* High-resolution de novo structure prediction from primary sequence. *bioRxiv* 2022.07.21.500999 (2022) doi:10.1101/2022.07.21.500999.

13. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1–11 (2022).

14. [No title]. *X (formerly Twitter)* https://twitter.com/Ag_smith/status/1417063635000598528?lang=en-GB.

15. [No title]. *X (formerly Twitter)* https://twitter.com/minkbaek/status/1417538291709071362?lang=en.

16.Qiao, Z., Nie, W., Vahdat, A., Miller, T. F., III & Anandkumar, A. State-specific protein-ligand complex structure prediction with a multi-scale deep generative model. arXiv preprint arXiv:2209.15171 (2022) doi:10.48550/arXiv.2209.15171.

17.Nakata, S., Mori, Y. & Tanaka, S. End-to-end protein–ligand complex structure generation with diffusion-based generative models. *BMC Bioinformatics* **24**, 1–18 (2023).

18.Baek, M. *et al.* Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods* 1–5 (2023).

19.Townshend, R. J. L. *et al.* Geometric deep learning of RNA structure. *Science* **373**, 1047–1051 (2021).

20.Jiang, D. *et al.* InteractionGraphNet: A Novel and Efficient Deep Graph Representation Learning Framework for Accurate Protein-Ligand Interaction Predictions. *J. Med. Chem.* **64**, (2021).

21.Jiang, H. *et al.* Predicting Protein–Ligand Docking Structure with Graph Neural Network. *J. Chem. Inf. Model.* (2022) doi:10.1021/acs.jcim.2c00127.

22.Corso, G., Stärk, H., Jing, B., Barzilay, R. & Jaakkola, T. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. arXiv preprint arXiv:2210.01776 (2022) doi:10.48550/arXiv.2210.01776.

23.Stärk, H., Ganea, O.-E., Pattanaik, L., Barzilay, R. & Jaakkola, T. EquiBind: Geometric Deep Learning for Drug Binding Structure Prediction. arXiv preprint arXiv:2202.05146 (2022) doi:10.48550/arXiv.2202.05146.

24.Liao, Z. et al. DeepDock: Enhancing Ligand-protein Interaction Prediction by a Combination of Ligand and Structure Information. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 311 – 317 (2019).

25.Lu, W. *et al.* TANKBind: Trigonometry-Aware Neural NetworKs for Drug-Protein Binding Structure Prediction. *Adv. Neural Inf. Process. Syst.* **35**, 7236–7249 (2022).

26.Zhou, G. *et al.* Uni-Mol: A Universal 3D Molecular Representation Learning Framework. (2023) doi:10.26434/chemrxiv-2022-jjm0j-v4.

27.Shen, T. *et al.* E2Efold-3D: End-to-End Deep Learning Method for accurate de novo RNA 3D

Structure Prediction. (2022).

28.van Dijk, M. & Bonvin, A. M. J. J. Pushing the limits of what is achievable in protein–DNA docking: benchmarking HADDOCK's performance. *Nucleic Acids Res.* **38**, 5634–5647 (2010).

29.Krishna, R. et al. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. Science 384, eadl2528 (2024).

30.Buttenschoen, M., Morris, G. M. & Deane, C. M. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. arXiv preprint arXiv:2308.05777 (2023) doi:10.48550/arXiv.2308.05777.

31.Das, R. *et al.* Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins* **91**, (2023).

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

33.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**, 2722–2728 (2013).

34.Karras, T., Aittala, M., Aila, T. & Laine, S. Elucidating the Design Space of Diffusion-Based Generative Models. arXiv preprint arXiv:2206.00364 (2022) doi:10.48550/arXiv.2206.00364.

35.Wang, Y., Elhag, A. A., Jaitly, N., Susskind, J. M. & Bautista, M. A. Generating Molecular Conformer Fields. arXiv preprint arXiv:2311.17932 (2023) doi:10.48550/arXiv.2311.17932.

36.Ji, Z. et al. Survey of Hallucination in Natural Language Generation. arXiv preprint arXiv:2202.03629 (2022) doi:10.1145/3571730.

37.Del Conte, A. *et al.* Critical assessment of protein intrinsic disorder prediction (CAID) - Results of round 2. *Proteins: Struct. Funct. Bioinf.* **91**, 1925–1934 (2023).

38.Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

39.Miller, E. B. *et al.* Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein–Ligand Binding. *J. Chem. Theory Comput.* (2021) doi:10.1021/acs.jctc.1c00136.

40.Baek, M. *et al.* Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods* 1–5 (2023).

41.Chen, K., Zhou, Y., Wang, S. & Xiong, P. RNA tertiary structure modeling with BRiQ potential in CASP15. *Proteins: Struct. Funct. Bioinf.* **91**, 1771–1778 (2023).

42.Basu, S. & Wallner, B. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLoS One* **11**, e0161879 (2016).

43.Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710 (2004).

44.Watson, E. R. *et al.* Molecular glue CELMoD compounds are regulators of cereblon conformation. *Science* **378**, 549–553 (2022).

45.Wayment-Steele, H. K. *et al.* Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* 1–3 (2023).

46.del Alamo, D., Sala, D., Mchaourab, H. S. & Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. (2022) doi:10.7554/eLife.75751.

47.Heo, L. & Feig, M. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins* **90**, 1873–1885 (2022).

48.Wallner, B. AFsample: improving multimer prediction with AlphaFold using massive sampling. *Bioinformatics* **39**, (2023).

Methods

Full algorithm details

Extensive explanations of the components are available in **Supplementary Methods 2–5** In addition, pseudocode is available in **Supplementary Algorithms 1–31**, network diagrams in **Fig. 1d, Fig. 2a,b,c, and Supplementary Fig. 2**, input features in **Supplementary Table 5,** and additional hyper parameters for training in **Supplementary Tables 3, 4, 7**.

Training regime

No structural data used during training was released after 2021-09-30, and for the model used in PoseBusters evaluations we filtered out PDB[32] structures released after 2019-09-30. One optimizer step uses a mini batch of 256 input data samples and during initial training 256 * 48 = 12,288 diffusion samples. For fine tuning the number of diffusion samples is reduced to 256 * 32 = 8,192. The model is trained in three stages, the initial training with a crop size of 384 tokens and two sequential fine tuning stages with crop sizes 640 and 768 tokens. See **Supplementary Methods 5.2** for more details**.**

Inference regime

No inference time templates or reference ligand position features were released after 2021-09-30, and in the case of PoseBusters evaluation, an earlier cutoff date of 2019-09-30 was used. The model can be run with different random seeds to generate alternative results, with a batch of diffusion samples per seed. Unless otherwise stated, all results are generated by selecting the top confidence sample from running 5 seeds of the same trained model, with 5 diffusion samples per model seed, for a total of 25 samples to choose from. Standard crystallisation aids are excluded from predictions (see **Supplementary Table 8**).

Results are shown for the top ranked sample and sample ranking depends on whether trying to select the overall best output globally, or the best output for some chain, interface or modified residue. Global ranking uses a mix of pTM and ipTM along with terms to reduce cases with large numbers of clashes and increase rates of disorder, individual chain ranking uses a chain specific pTM measure, interface ranking uses a bespoke ipTM measure for the relevant chain pair and modified residue ranking uses average pLDDT over the residue of interest (see **Supplementary Methods 5.9.3** for details).

Metrics

Evaluation compares a predicted structure to the corresponding ground truth structure. If the complex contains multiple identical entities, assignment of the predicted units to the ground truth units is found by maximising LDDT. Assignment in local symmetry groups of atoms in ligands is solved by exhaustive search over the first 1000 per-residue symmetries as given by RDKit.

We measure the quality of the predictions with DockQ, LDDT or pocket-aligned RMSD. For nucleic-protein interfaces we measure interface accuracy via interface LDDT (iLDDT), which is calculated from distances between atoms across different chains in the interface. DockQ and iLDDT are highly correlated (**Extended Data Fig. 9**), so the standard cutoffs for DockQ can be translated to equivalent iLDDT cutoffs. Nucleic acid LDDTs (intra-chains and interface) were calculated with an inclusion radius of 30 Å compared to the usual 15 Å used for proteins, owing to their larger scale. For confidence calibration assessment, we use a bespoke LDDT, "LDDT_to_polymer" metric which considers differences from each atom of a given entity to any

C^{α} or C1' polymer atom within its inclusion radius. This is closely related to how the confidence prediction is trained (see **Supplementary Methods 4.3.1** for details).

Pocket-aligned RMSD is computed as follows: the pocket is defined as all heavy atoms within 10 Å of any heavy atom of the ligand, restricted to the primary polymer chain for the ligand or modified residue being scored, and further restricted to only backbone atoms for proteins. The primary polymer chain is defined variously: for PoseBusters it is the protein chain with the most atoms within 10 Å of the ligand, for bonded ligand scores it is the bonded polymer chain and for modified residues it is the chain that the residue is contained in (minus that residue). The pocket is used to align the predicted structure to the ground truth structure with least squares rigid alignment and then RMSD is computed on all heavy atoms of the ligand.

Recent PDB evaluation set

General model evaluation was performed on our Recent PDB set consisting of 8,856 PDB complexes released between 2022-05-01 and 2023-01-12. The set contains almost all PDB complexes released during that period less than 5,120 model tokens in size (see **Supplementary Methods 6.1** for details). Single chains and interfaces within each structure were scored separately rather than only looking at full complex scores, then clustering was applied to chains and interfaces so that scores could be aggregated first within clusters and then across clusters for mean scores, or using a weighting of inverse cluster size for distributional statistics (see **Supplementary Methods 6.2 and 6.4** for details).

Evaluation on ligands excludes standard crystallisation aids (**Supplementary Table 8**), our ligand exclusion list (**Supplementary Table 9**) and glycans (**Supplementary Table 10**). Bonded and non-bonded ligands are evaluated separately. Ions are only included when specifically mentioned (see **Supplementary Table 11**).

The Recent PDB set is filtered to a low homology subset (see **Supplementary Methods 6.1**) for some results where stated. Homology is defined as sequence identity to sequences in the training set and is measured via template search (see **Supplementary Methods 2.4** for details). Individual polymer chains in evaluation complexes are filtered out if the maximum sequence identity to chains in the training set is greater than 40%, where sequence identity is the percent of residues in the evaluation set chain that are identical to the training set chain. Individual peptide chains (protein chains with less than 16 residues) are always filtered out. For polymer-polymer interfaces, if both polymers have greater than 40% sequence identity to two chains in the same complex in the training set, then the interface is filtered out. For interfaces to a peptide the interface is filtered out if the non-peptide entity has greater than 40% sequence identity to any chain in the training set.

To compare quality of prediction of protein-protein interfaces and protein monomers against that of AlphaFold-Multimer v2.3 (AF-M 2.3[8]), and to compare dependence of single protein chain prediction quality on MSA depth, we restrict the low homology Recent PDB set to complexes

with fewer than 20 protein chains and fewer than 2,560 tokens. We compare against unrelaxed AF-M 2.3 predictions.

To study antibody-antigen interface prediction, we filter the low homology Recent PDB set to complexes that contain at least one protein-protein interface where one of the protein chains is in one of the two largest PDB chain clusters (these clusters are representative of antibodies). We further filter to complexes with at most 2,560 tokens and with no unknown amino acids in PDB, to allow extensive comparison against relaxed predictions of AlphaFold-Multimer v2.3. That leaves 71 antibody-antigen complexes, containing 166 antibody-antigen interfaces spanning 65 interface clusters.

MSA depth analysis **(Extended Data Fig. 7a)** was based on computing the normalised number of effective sequences (N_{eff}) for each position of a query sequence. Per-residue N_{eff} values were obtained by counting the number of non-gap residues in the MSA for this position and weighting the sequences using the N_{eff} scheme[49] with a threshold of 80% sequence identity measured on the region that is non-gap in either sequence.

Nucleic acid prediction baseline

For benchmarking performance on nucleic acid structure prediction, we report baseline comparisons to an existing machine learning system for protein-nucleic acid and RNA tertiary structure prediction, RoseTTAFold2NA[18]. We run the open source RF2NA[50] with the same multiple sequence alignments (MSAs) as were used for AlphaFold 3 predictions. For comparison between AlphaFold 3 and RF2NA, a subset of our Recent PDB set are chosen to meet the RF2NA criteria (<1000 total residues and nucleotides). As RF2NA was not trained to predict systems with DNA and RNA, analysis is limited to targets with only one nucleic acid type. No system was publically available at time of writing for baseline comparisons on data with arbitrary combinations of biomolecular types in PDB.

As an additional baseline for RNA tertiary structure prediction, we evaluate AlphaFold 3 performance on CASP15 RNA targets that are currently publicly available (R1116/8S95, R1117/8FZA, R1126 (downloaded from the CASP 15 website https://predictioncenter.org/casp15/TARGETS_PDB/R1126.pdb), R1128/8BTZ, R1136/7ZJ4, R1138/[7PTK/7PTL], R1189/7YR7, and R1190/7YR6). We compare top-1 ranked predictions, and where multiple ground truth structures exist (R1136) the prediction is scored against the closest state. We display comparisons to RF2NA as a representative machine learning system, AIchemy_RNA2 as the top performing entrant with human intervention, and AIchemy_RNA as the top performing machine learning system. All entrants' predictions were downloaded from the CASP website and scored internally.

PoseBusters

While other analyses used an AlphaFold model trained on PDB data released prior to a cutoff of 2021-09-30, our PoseBusters analysis was conducted on a model (with identical architecture and similar training schedule) differing only in the use of an earlier 2019-09-30 cutoff. This analysis therefore did not include training data, inference time templates, or "ref_pos" features released after this date.

Inference was performed on the asymmetric unit from specified PDBs, with the following minor modifications. In several PDB files, chains clashing with the ligand of interest were removed (7O1T, 7PUV, 7SCW, 7WJB, 7ZXV, 8AIE). Another PDB (8F4J) was too large to inference the entire system (over 5120 tokens), so we only included protein chains within 20 Å of the ligand of interest. Five model seeds, each with five diffusion samples, were produced per target, resulting in 25 predictions, which were ranked by quality and predicted accuracy: the ranking score was calculated from an ipTM aggregate (**Supplementary Methods 5.9.3** point 3), then further divided by 100 if the ligand had chirality errors or had clashes with the protein.

For pocket-aligned RMSD, first alignment between the predicted and ground truth structures was conducted by aligning to ground truth pocket backbone atoms (CA, C, or N atoms within 10 Å of the ligand of interest) from the primary protein chain (the chain with the greatest number of contacts within 10 Å of the ligand). The posebusters python package v0.2.7[51] was used to score RMSD and violations from the pocket-aligned predictions.

While AlphaFold models are "blind" to the protein pocket, docking is often performed with knowledge of the protein pocket residues. For example, Uni-Mol specifies the pocket as any residue within 6 Å of the heavy atoms in the ligand of interest[26]. To evaluate the ability of AlphaFold 3 to "dock" ligands accurately when given pocket information, we fine-tuned a 2019-09-30-cutoff AlphaFold 3 model with an additional token feature specifying pocket-ligand pairs (**Supplementary Methods 2.8**). Specifically, an additional token feature was introduced, set to true for a ligand entity of interest and any pocket residues with heavy atoms within 6 Å of the ligand entity. At training time a single random ligand entity is chosen to use in this feature. Note that multiple ligand chains with the same entity (CCD code) may be selected. At inference time, the ligand entity was chosen based on the ligand of interest's CCD code, so again multiple ligand chains were occasionally chosen. Results of this analysis are shown in **Extended Data Fig. 4**.

Model Performance Analysis and Visualization

Data analysis used Python v3.11.7 (https://www.python.org/), NumPy v1.26.3 (https://github.com/numpy/numpy), SciPy v1.9.3 (https:// www.scipy.org/), seaborn v0.12.2 (https://github.com/mwaskom/seaborn), Matplotlib v3.6.1 (https://github.com/matplotlib/matplotlib), pandas v2.0.3 (https://github.com/pandas-dev/pandas), statsmodels v0.12.2 (https://github.com/statsmodels/statsmodels), RDKit v4.3.0 (https://github.com/rdkit/rdkit), and Colab (https://research.google.com/colaboratory). TM-align

v20190822 (https://zhanglab.dcmb.med.umich.edu/TM-align/) was used for computing TM-scores. Structure visualizations were created in Pymol v2.55.5 (https://github.com/schrodinger/pymol-open-source).

Data availability

All scientific datasets used to create training and evaluation inputs are freely available from public sources. Structures from the PDB were used for training and as templates (https://files.wwpdb.org/pub/pdb/data/assemblies/mmCIF/; for sequence clusters see https://cdn.rcsb.org/resources/sequence/clusters/clusters-by-entity-40.txt; for sequence data see https://files.wwpdb.org/pub/pdb/derived_data/). Training used a version of the PDB downloaded 12 January 2023, while template search used a version downloaded 28 September 2022. We also used the Chemical Components Dictionary downloaded on 19 October 2023 (https://www.wwpdb.org/data/ccd).
We show experimental structures from the PDB with accession numbers 7PZB[52,53], 7PNM[54,55], 7TQL[56,57], 7AU2[58,59], 7U8C[60,61], 7URD[62,63], 7WUX[64,65], 7QIE[66,67], 7T82[68,69], 7CTM[70,71], 8CVP[44,72], 8D7U[44,73], 7F60[74,75], 8BTI[76,77], 7KZ9[78,79], 7XFA[80,81], 7PEU[82,83], 7SDW[84,85], 7TNZ[86,87], 7R6R [88,89], 7USR[90,91], and 7Z1K.[92,93]

We also used the following publicly available databases for training or evaluation. Detailed usage is described in **Supplementary Methods 2.2** and **Supplementary Methods 2.5.2**.
UniRef90 v.2020_01 (https://ftp.ebi.ac.uk/pub/databases/uniprot/previous_releases/release-2020_01/uniref/),
UniRef90 v.2020_03 (https://ftp.ebi.ac.uk/pub/databases/uniprot/previous_releases/release-2020_03/uniref/),
UniRef90 v.2022_05 https://ftp.ebi.ac.uk/pub/databases/uniprot/previous_releases/release-2022_05/uniref/),
Uniclust30 v.2018_08 (https://wwwuser.gwdg.de/~compbiol/uniclust/2018_08/),
Uniclust30 v.2021_03 (https://wwwuser.gwdg.de/~compbiol/uniclust/2021_03/),
MGnify clusters v.2018_12 (https://ftp.ebi.ac.uk/pub/databases/metagenomics/peptide_database/2018_12/),
MGnify clusters v.2022_05 (https://ftp.ebi.ac.uk/pub/databases/metagenomics/peptide_database/2022_05/),
BFD (https://bfd.mmseqs.com),
RFam v.14.9 (https://ftp.ebi.ac.uk/pub/databases/Rfam/14.9/),
RNAcentral v.21.0

(https://ftp.ebi.ac.uk/pub/databases/RNAcentral/releases/21.0/),
Nucleotide Database (as of 23 February 2023)
(https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nt.gz),
JASPAR 2022
(https://jaspar.elixir.no/downloads/; see https://jaspar.elixir.no/profile-versions for version information),
SELEX protein sequences from Supplementary Tables[94]
(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8009048/),
SELEX protein sequences from Supplementary Tables[95]
(https://www.nature.com/articles/nature15518).

Code availability

AlphaFold 3 will be available as a non-commercial usage only server at https://www.alphafoldserver.com, with restrictions on allowed ligands and covalent modifications. Pseudocode describing the algorithms is available in the Supplementary Information. Code is not provided.

Acknowledgements

We thank Giorgio Arena, Žiga Avsec, Anthony Baryshnikov, Russ Bates, Molly Beck, Aliyah Bond, Nathalie Bradley-Schmieg, Jana Cavojska, Ben Coppin, Emilien Dupont, Sean Eddy, Marco Fiscato, Richard Green, Dhavanthi Hariharan, Kristian Holsheimer, Nicole Hurley, Chris Jones, Koray Kavukcuoglu, Jacob Kelly, Eugenia Kim, Anna Koivuniemi, Oleg Kovalevskiy, Dariusz Lasecki, Meera Last, Agata Laydon, William McCorkindale, Sam Miller, Alex Morris, Lauren Nicolaisen, Evan Palmer, Antonia Paterson, Stig Petersen, Ollie Purkiss, Chongyang Shi, George Thomas, Gregory Thornton and Hamish Tomlinson for their contributions

Author Contributions

The equally contributing authors are alphabetically ordered, as are the remaining core contributor authors (excluding jointly supervising authors) and similar for all remaining non-supervising authors. D.H., M.J. and J.J. led the research. M.J., J.J. and P.K. developed research strategy. J. Abramson, V.B., T.G. and C.-C.H. led key research pillars. T.G. and A. Žídek led the technical framework for research. O.B., H.G. and S.S. coordinated and managed the Research Project. J. Abramson, J. Adler, E.A., A. Ballard, J.B., V.B., A.C.-R., J.D., R.E., D.A.E., M.F., F.F., T.G., C.-C.H., M.J., J.J., Y.A.K., A. Potapenko, A. Pritzel, D.R., O.R., A.T., C.T., K.T., L.W., Z.W. and E.Z. developed the neural network architecture and training procedure. J. Abramson, A. Ballard, J.B., V.B., C.B., S.B., A. Bridgland, A. Cherepanov, A.C.-R., A. Cowie, J.D., T.G., R.J., M.O., K.P., D.R., O.R., M.Z., A. Žemgulytė and A. Žídek developed the training, inference, data, and evaluation infrastructure. J. Abramson, J. Adler, A. Ballard, V.B., A.C.-R., R.E.,

D.A.E., T.G., D.H., M.J., J.J., P.K., K.P., A. Pritzel, O.R., P.S., S.S., A.S., K.T. and L.W. contributed to the writing of the paper. M.C., C.M.R.L., S.Y. advised on the project.

Competing Interests

Author-affiliated entities have filed US provisional patent applications including 63/611,674, 63/611,638 and 63/546,444 relating to predicting three-dimensional (3d) structures of molecule complexes using embedding neural networks and generative models. All authors other than A. Bridgland, Y.A.K. and E.Z. have commercial interests in the work described.

Further Citations

49.Wu, T., Hou, J., Adhikari, B. & Cheng, J. Analysis of several key factors influencing deep learning-based inter-residue contact prediction. *Bioinformatics* **36**, 1091–1098 (2020).

50.Release RF2NA v0.2 · uw-ipd/RoseTTAFold2NA. *GitHub* https://github.com/uw-ipd/RoseTTAFold2NA/releases/tag/v0.2.

51.Release v0.2.7 · maabuu/posebusters. *GitHub* https://github.com/maabuu/posebusters/releases/tag/v0.2.7.

52.Werel, L. *et al.* Structural Basis of Dual Specificity of Sinorhizobium meliloti Clr, a cAMP and cGMP Receptor Protein. *MBio* **14**, e0302822 (2023).

53.wwPDB: 7PZB. https://doi.org/10.2210/pdb7PZB/pdb.

54.Wang, C. *et al.* Antigenic structure of the human coronavirus OC43 spike reveals exposed and occluded neutralizing epitopes. *Nat. Commun.* **13**, 1–15 (2022).

55.wwPDB: 7PNM. https://doi.org/10.2210/pdb7PNM/pdb.

56.Lapointe, C. P. *et al.* eIF5B and eIF1A reorient initiator tRNA to allow ribosomal subunit joining. *Nature* **607**, 185–190 (2022).

57.wwPDB: 7TQL. https://doi.org/10.2210/pdb7TQL/pdb.

58.Wilson, L. F. L. *et al.* The structure of EXTL3 helps to explain the different roles of bi-domain exostosins in heparan sulfate synthesis. *Nat. Commun.* **13**, 1–15 (2022).

59.wwPDB: 7AU2. https://doi.org/10.2210/pdb7AU2/pdb.

60.Liu, X. *et al.* Highly active CAR T cells that bind to a juxtamembrane region of mesothelin and

are not blocked by shed mesothelin. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2202439119 (2022).

61.wwPDB: 7U8C. https://doi.org/10.2210/pdb7U8C/pdb.

62.Liu, Y. *et al.* Mechanisms and inhibition of Porcupine-mediated Wnt acylation. *Nature* **607**, 816–822 (2022).

63.wwPDB: 7URD. https://doi.org/10.2210/pdb7URD/pdb.

64.Kurosawa, S. *et al.* Molecular Basis for Enzymatic Aziridine Formation via Sulfate Elimination. *J. Am. Chem. Soc.* **144**, 16164–16170 (2022).

65.wwPDB: 7WUX. https://doi.org/10.2210/pdb7WUX/pdb.

66.Boffey, H. K. *et al.* Development of Selective Phosphatidylinositol 5-Phosphate 4-Kinase γ Inhibitors with a Non-ATP-competitive, Allosteric Binding Mode. *J. Med. Chem.* **65**, 3359–3370 (2022).

67.wwPDB: 7QIE. https://doi.org/10.2210/pdb7QIE/pdb.

68.Buckley, P. T. *et al.* Multivalent human antibody-centyrin fusion protein to prevent and treat Staphylococcus aureus infections. *Cell Host Microbe* **31**, 751–765.e11 (2023).

69.wwPDB: 7T82. https://doi.org/10.2210/pdb7T82/pdb.

70.Mohapatra, S. B. & Manoj, N. Structural basis of catalysis and substrate recognition by the NAD(H)-dependent α-d-glucuronidase from the glycoside hydrolase family 4. *Biochem. J* **478**, 943–959 (2021).

71.wwPDB: 7CTM. https://doi.org/10.2210/pdb7CTM/pdb.

72.wwPDB: 8CVP. https://doi.org/10.2210/pdb8CVP/pdb.

73.wwPDB: 8D7U. https://doi.org/10.2210/pdb8D7U/pdb.

74.Gao, X. *et al.* Structural basis for Sarbecovirus ORF6 mediated blockage of nucleocytoplasmic transport. *Nat. Commun.* **13**, 1–11 (2022).

75.wwPDB: 7F60. https://doi.org/10.2210/pdb7F60/pdb.

76.Atkinson, B. N. *et al.* Designed switch from covalent to non-covalent inhibitors of carboxylesterase Notum activity. *Eur. J. Med. Chem.* **251**, 115132 (2023).

77.wwPDB: 8BTI. https://doi.org/10.2210/pdb8BTI/pdb.

78.Luo, S. *et al.* Structural basis for a bacterial Pip system plant effector recognition protein. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).

79.wwPDB: 7KZ9. https://doi.org/10.2210/pdb7KZ9/pdb.

80.Liu, C. *et al.* Identification of Monosaccharide Derivatives as Potent, Selective, and Orally Bioavailable Inhibitors of Human and Mouse Galectin-3. *J. Med. Chem.* **65**, 11084–11099 (2022).

81.wwPDB: 7XFA. https://doi.org/10.2210/pdb7XFA/pdb.

82.Dombrowski, M., Engeholm, M., Dienemann, C., Dodonova, S. & Cramer, P. Histone H1 binding to nucleosome arrays depends on linker DNA length and trajectory. *Nat. Struct. Mol. Biol.* **29**, 493–501 (2022).

83.wwPDB: 7PEU. https://doi.org/10.2210/pdb7PEU/pdb.

84.Vecchioni, S. *et al.* Metal-Mediated DNA Nanotechnology in 3D: Structural Library by Templated Diffraction. *Adv. Mater.* **35**, e2210938 (2023).

85.wwPDB: 7SDW. https://doi.org/10.2210/pdb7SDW/pdb.

86.Wang, W. & Pyle, A. M. The RIG-I receptor adopts two different conformations for distinguishing host from viral RNA ligands. *Mol. Cell* **82**, 4131–4144.e6 (2022).

87.wwPDB: 7TNZ. https://doi.org/10.2210/pdb7TNZ/pdb.

88.McGinnis, R. J. *et al.* A monomeric mycobacteriophage immunity repressor utilizes two domains to recognize an asymmetric DNA sequence. *Nat. Commun.* **13**, 4105 (2022).

89.wwPDB: 7R6R. https://doi.org/10.2210/pdb7R6R/pdb.

90.Dietrich, M. H. *et al.* Nanobodies against Pfs230 block Plasmodium falciparum transmission. *Biochem. J* **479**, 2529–2546 (2022).

91.wwPDB: 7USR. https://doi.org/10.2210/pdb7USR/pdb.

92.Appel, L.-M. *et al.* The SPOC domain is a phosphoserine binding module that bridges transcription machinery with co- and post-transcriptional regulators. *Nat. Commun.* **14**, 1–22 (2023).

93.wwPDB: 7Z1K. https://doi.org/10.2210/pdb7Z1K/pdb.

94.Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, (2017).

95.Jolma, A. *et al.* DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527**, 384–388 (2015).

Extended Data Figure Captions

Extended Data Figure 1 | Disordered region prediction. a, Example prediction for a disordered protein from AlphaFoldMultimer v2.3, AlphaFold 3, and AlphaFold 3 trained without the disordered protein PDB cross distillation set. Protein is DP02376 from the CAID 2 (Critical Assessment of protein Intrinsic Disorder prediction) set. Predictions coloured by pLDDT (orange: pLDDT<=50 ,yellow: 50<pLDDT<=70, light blue: 70<pLDDT<=90, and dark blue: 90<=pLDDT<100). **b**, Predictions of disorder across residues in proteins in the CAID 2 set, which are also low homology to the AF3 training set. Prediction methods include RASA (relative accessible surface area) and pLDDT (N=151 proteins; 46,093 residues).

Extended Data Figure 2 | Accuracy across training. Training curves for initial training and fine tuning showing LDDT (local distance difference test) on our evaluation set as a function of optimizer steps. One optimizer step uses a mini batch of 256 trunk samples and during initial training 256 * 48 = 12,288 diffusion samples. For fine tuning the number of diffusion samples is reduced to 256 * 32 = 8,192. The scatter plot shows the raw data points and the lines show the smoothed performance using a median filter with a kernel width of 9 data points. The dashed lines mark the points where the smoothed performance passes 90% and 97% of the initial training maximum for the first time.

Extended Data Figure 3 | AlphaFold 3 predictions of PoseBusters examples for which Vina and Gold were inaccurate. Predicted protein chains are shown in blue, predicted ligands in orange, and ground truth in grey. **a**, Human Notum bound to inhibitor ARUK3004556 (PDB ID 8BTI, ligand RMSD: 0.65 Å). **b**, Pseudomonas sp. PDC86 Aapf bound to HEHEAA (PDB ID 7KZ9, ligand RMSD: 1.3 Å). **c**, Human Galectin-3 carbohydrate-recognition domain in complex with compound 22 (PDB ID 7XFA, ligand RMSD: 0.44 Å).

Extended Data Figure 4 | PoseBusters analysis. a, Comparison of AlphaFold 3 and baseline method protein-ligand binding success on the PoseBusters Version 1 benchmark set (V1, August 2023 release). Methods classified by the extent of ground truth information used to make predictions. Note all methods that use pocket residue information except for UMol and AF3 also use ground truth holo protein structures. **b**, PoseBusters Version 2 (V2, November 2023 release) comparison between the leading docking method Vina and AF3 2019 (two-sided Fisher exact test, N = 308 targets, $p = 2.3 * 10^{-8}$). **c**, PoseBusters V2 results of AF3 2019 on targets with low, moderate, and high protein sequence homology (integer ranges indicate maximum sequence identity with proteins in the training set). **d**, PoseBusters V2 results of AF3 2019 with ligands split by those characterised as “common natural” ligands and others. “Common natural” ligands are defined as those which occur greater than 100 times in the PDB and which are not non-natural (by visual inspection). A full list may be found in Supplementary Table 15. Dark bar indicates RMSD < 2 Å and passing PoseBusters validity checks (PB-valid). **e**, PoseBusters V2 structural accuracy and validity. Dark bar indicates RMSD < 2 Å and passing PoseBusters validity checks (PB-valid). Light hashed bar indicates RMSD < 2 Å but not PB valid. **f**, PoseBusters V2 detailed validity check comparison. Error bars indicate exact binomial distribution 95% confidence intervals. N=427 targets for RoseTTAFold All-Atom and 428 targets for all others in Version 1; 308 targets in Version 2.

Extended Data Figure 5 | Nucleic acid prediction accuracy and confidences. a, CASP15 RNA prediction accuracy from AIChemy_RNA (the top AI-based submission), RoseTTAFold2NA (the AI-based method capable of predicting proteinRNA complexes), and AlphaFold 3. Ten of the 13 targets are available in the PDB or via the CASP15 website for evaluation. Predictions are downloaded from the CASP website for external models. **b**,

Accuracy on structures containing low homology RNA-only or DNA-only complexes from the recent PDB evaluation set. Comparison between AlphaFold 3 and RoseTTAFold2NA (RF2NA) (RNA: N=29 structures, paired Wilcoxon signed-rank test, p=1.6 * 10^{-7}; DNA: N=63 structures, paired two-sided Wilcoxon signed-rank test, p=5.2 * 10^{-12}). Note RF2NA was only trained and evaluated on duplexes (chains forming at least 10 hydrogen bonds), but some DNA structures in this set may not be duplexes. Box, centerline, and whiskers boundaries are at (25%, 75%) intervals, median, and (5%, 95%) intervals. **c** Predicted structure of a mycobacteriophage immunity repressor protein bound to double stranded DNA (PDB ID 7R6R), coloured by pLDDT (left; orange: 0-50, yellow: 50-70, cyan 70-90, and blue 90-100) and chain id (right). Note the disordered N-terminus not entirely shown. **d**, Predicted aligned error (PAE) per token-pair for the prediction in c with rows and columns labelled by chain id and green gradient indicating PAE.

Extended Data Figure 6 | Analysis and examples for modified proteins and nucleic acids. a, Accuracy on structures
containing common phosphorylation residues (SEP, TPO, PTR, NEP, HIP) from the recent PDB evaluation set. Comparison between AlphaFold 3 with phosphorylation modelled, and AlphaFold 3 without modelling phosphorylation (N=76 clusters, paired two-sided Wilcoxon signed-rank test, p=1.6 * 10^{-4}). Note, to predict a structure without modelling phosphorylation, we predict the parent (standard) residue in place of the modification. AlphaFold 3 generally achieves better backbone accuracy when modelling phosphorylation. Error bars indicate exact binomial distribution 95% confidence intervals. **b**, SPOC domain of human SHARP in complex with phosphorylated RNA polymerase II C-terminal domain (PDB ID 7Z1K), predictions coloured by pLDDT (orange: 0-50, yellow: 50-70, cyan 70-90, and blue 90-100). Left: Phosphorylation modelled (mean pocket-aligned $RMSD_{C\alpha}$ 2.104 Å). Right: Without modelling phosphorylation (mean pocketaligned $RMSD_{C\alpha}$ 10.261 Å). When excluding phosphorylation, AlphaFold 3 provides lower pLDDT confidence on the phosphopeptide. **c**, Structure of parkin bound to two phospho-ubiquitin molecules (PDB ID 7US1), predictions similarly coloured by pLDDT. Left: Phosphorylation modelled (mean pocket-aligned $RMSD_{C\alpha}$ 0.424 Å). Right: Without modelling phosphorylation (mean pocket-aligned $RMSD_{C\alpha}$ 9.706 Å). When excluding phosphorylation, AlphaFold 3 provides lower pLDDT confidence on the interface residues of the incorrectly predicted ubiquitin. **d**, Example structures with modified nucleic acids. Left: Guanosine monophosphate in RNA (PDB ID 7TNZ, mean pocket-aligned modified residue RMSD 0.840 Å). Right: Methylated DNA cytosines (PDB ID 7SDW, mean pocket-aligned modified residue RMSD 0.502 Å). Welabel residues of the predicted structure for reference. Ground truth structure in grey; predicted protein in blue, predicted RNA in purple, predicted DNA in magenta, predicted ions in orange, with predicted modifications highlighted via spheres

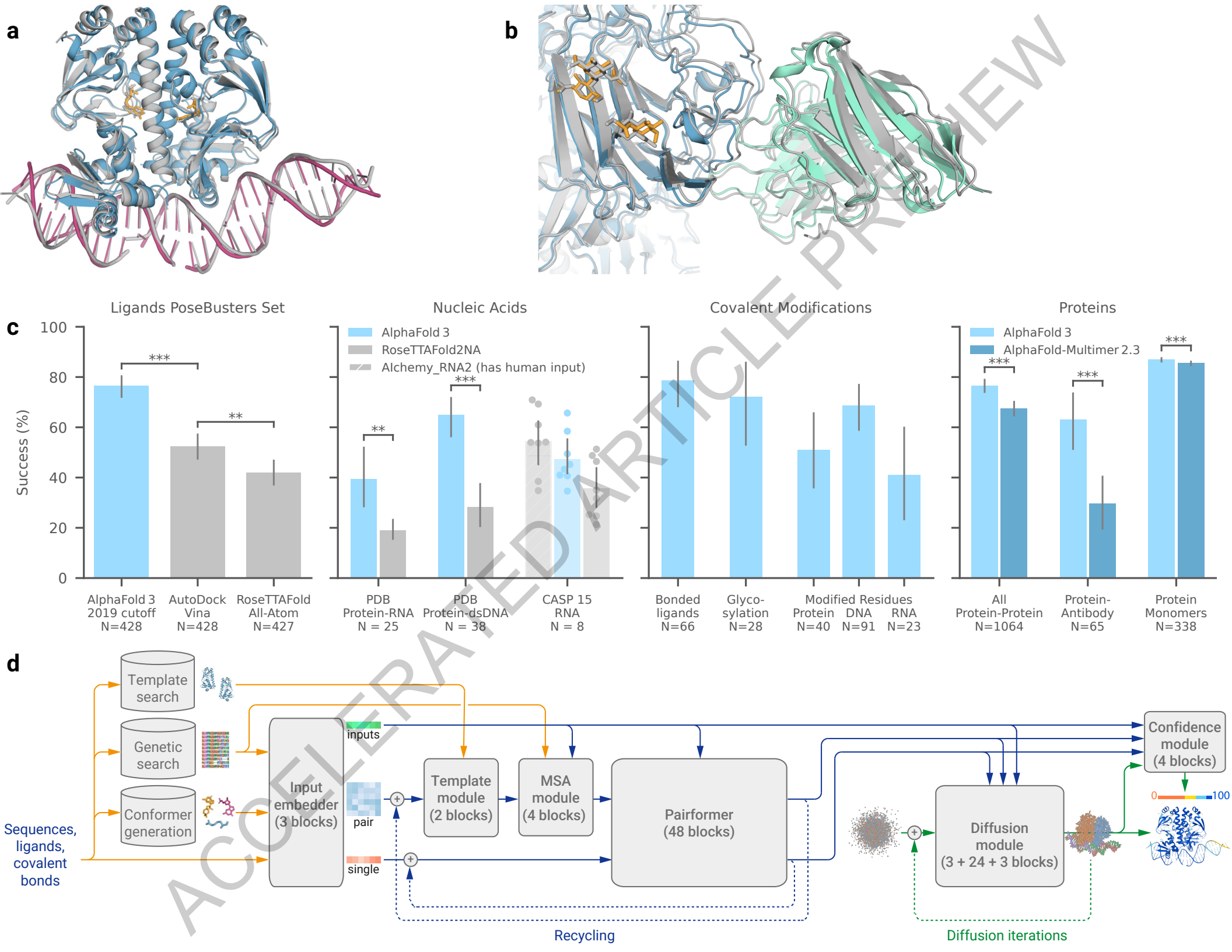
Extended Data Figure 7 | Model accuracy with MSA size and number of seeds. a, Effect of MSA depth on protein prediction accuracy. Accuracy is given as single chain LDDT score and MSA depth is computed by counting the number of non-gap residues for each position in the MSA using the N_{eff} weighting scheme and taking the median across residues (see Methods for details on N_{eff}). MSA used for AF-M 2.3 differs slightly from AF3; the data uses the AF3 MSA depth for both to make the comparison clearer. The analysis uses every protein chain in the low homology Recent PDB set, restricted to chains in complexes with fewer than 20 protein chains and fewer than 2,560 tokens (see Methods for details on Recent PDB set and comparisons to AF-M 2.3). The curves are obtained through Gaussian kernel average smoothing (window size is 0.2 units in $\log10(N_{eff})$); the shaded area is the 95% confidence interval estimated using bootstrap of 10,000 samples. **b,** Increase in ranked accuracy with number of seeds for different molecule types. Predictions are ranked by confidence, and only the most confident per interface is scored. Evaluated on the low homology recent PDB set, filtered to less than 1,536 tokens. Number of clusters evaluated: dna-intra=386, protein-intra=875, rnaintra=78, protein-dna=307, protein-rna=102, protein-protein (antibody=False)=697, protein-protein (antibody=True)=58. Confidence intervals are 95% bootstraps over 1,000 samples.

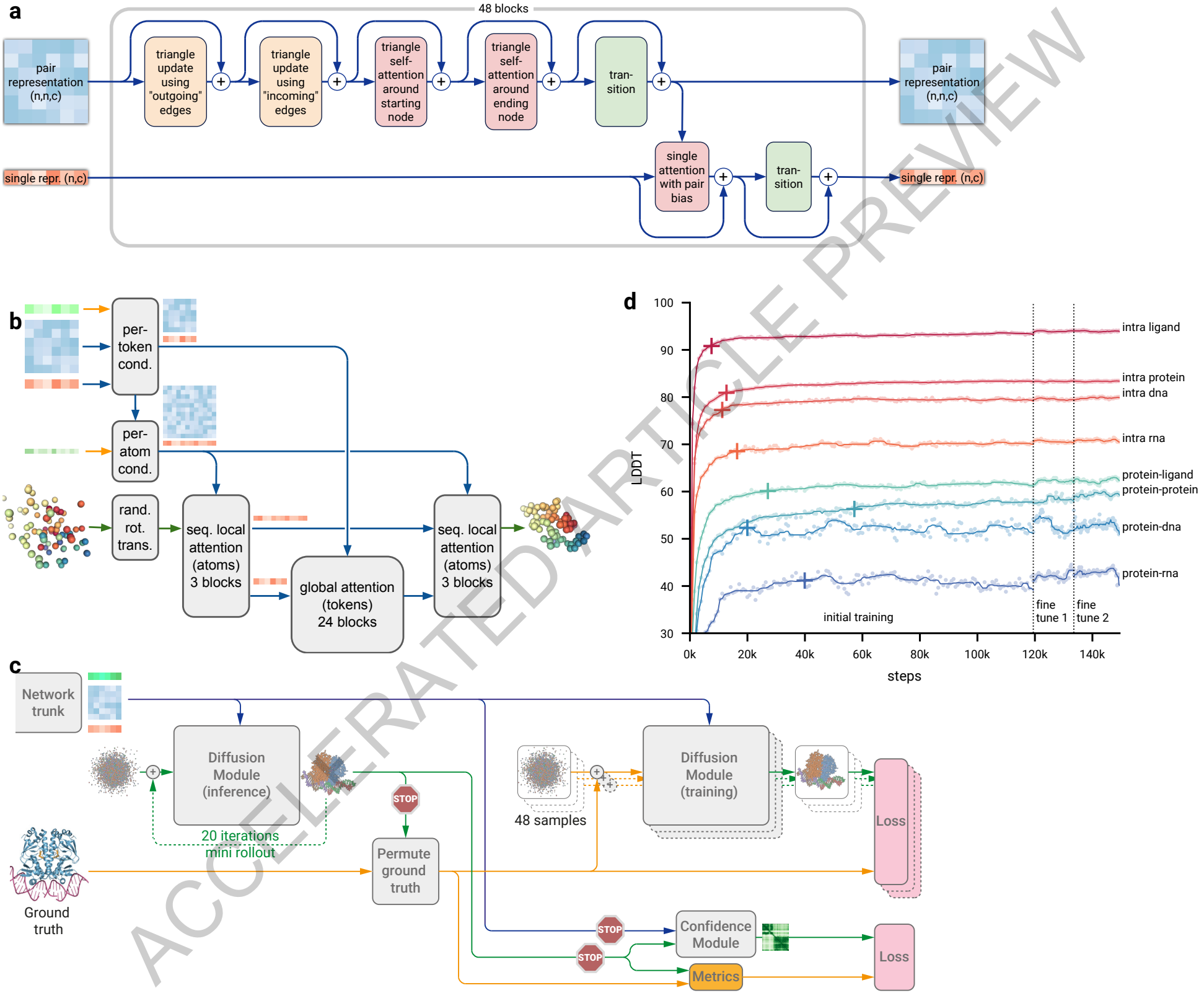
Extended Data Figure 8 | Relationship between confidence and accuracy for protein interactions with ions, bonded ligands and bonded glycans. Accuracy is given as the percentage of interface clusters under various pocket-aligned RMSD thresholds, as a function of the chain pair ipTM of the interface. The ions group includes both

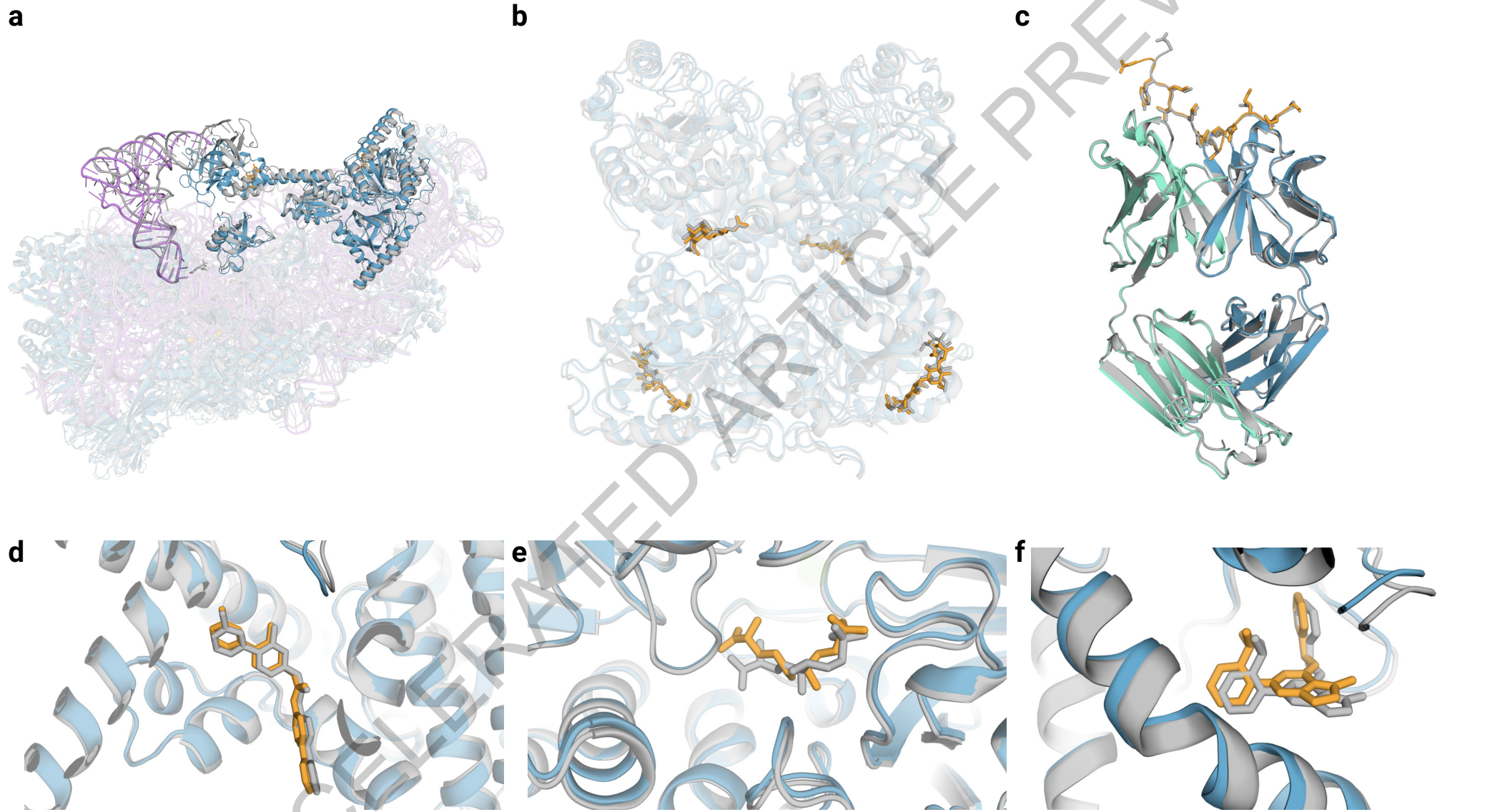
metals and nonmetals. N values report the number of clusters in each band. For a similar analysis on general ligand-protein interfaces, see Figure 4 of main text.

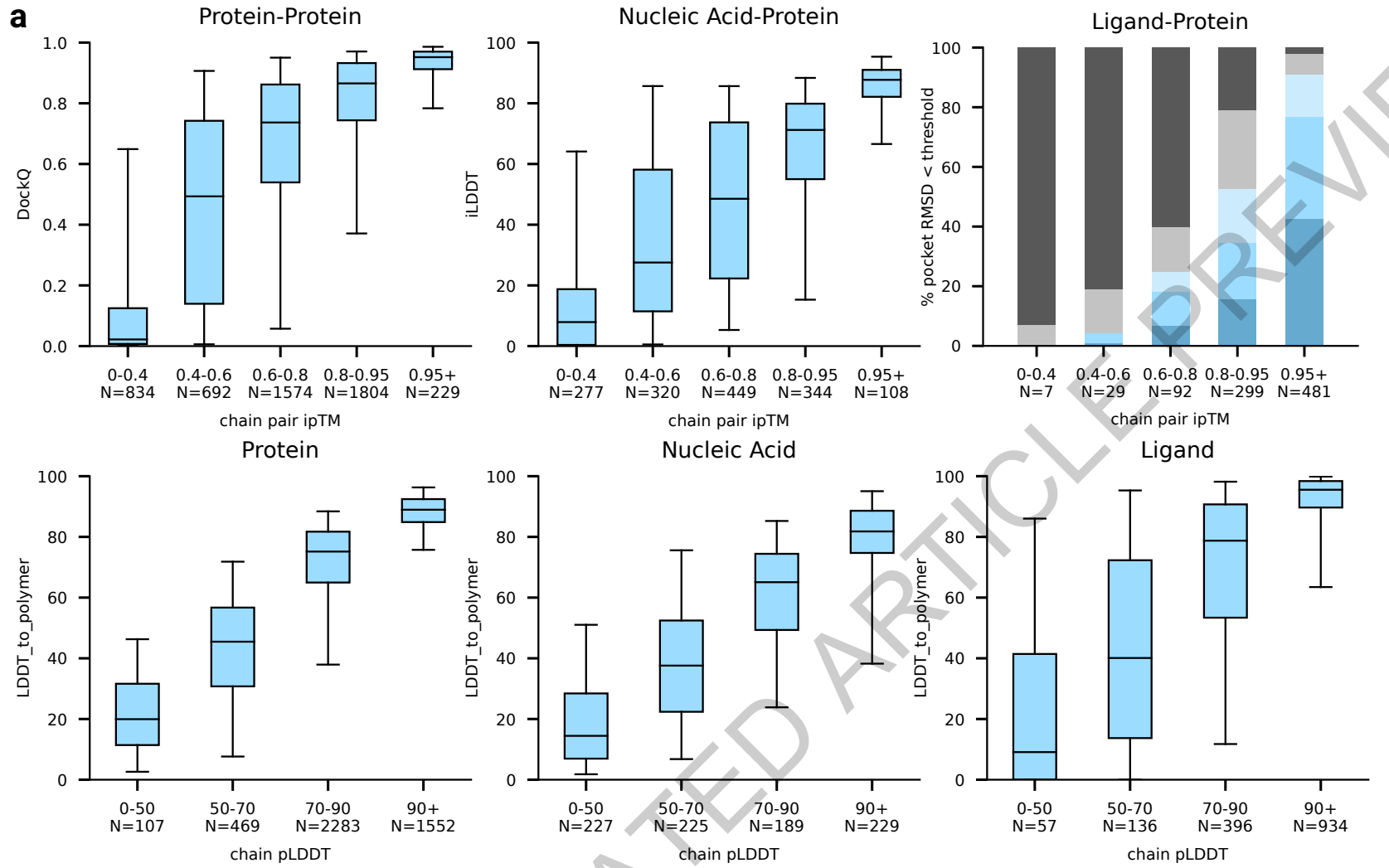
Extended Data Figure 9 |Correlation of DockQ and iLDDT for protein-protein interfaces. One data point per cluster, 4,182 clusters shown. Line of best fit with a Huber regressor with epsilon 1. DockQ categories correct (>0.23), and very high accuracy (>0.8) correspond to iLDDTs of 23.6 and 77.6 respectively

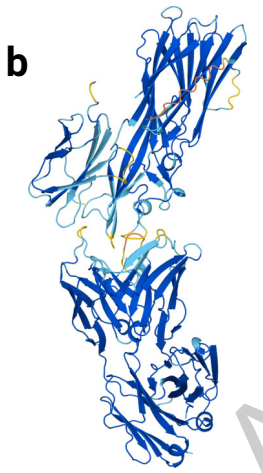
Extended Data Table 1 | Prediction accuracy across biomolecular complexes. AlphaFold 3 Performance on PoseBusters V1 (August 2023 release), PoseBusters V2 (November 6th 2023 release), and our Recent PDB evaluation set. For ligands and nucleic acids N indicates number of structures; for covalent modifications and proteins N indicates number of clusters.

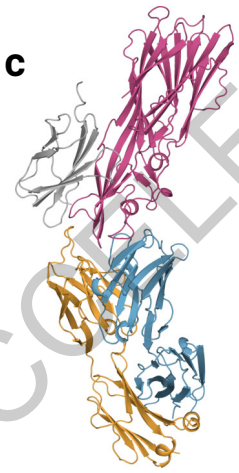


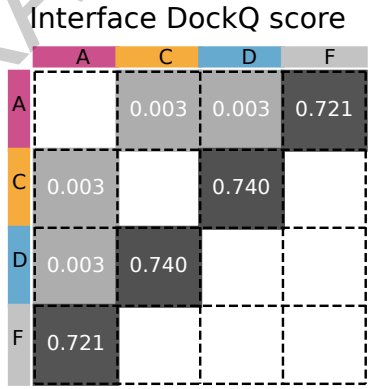


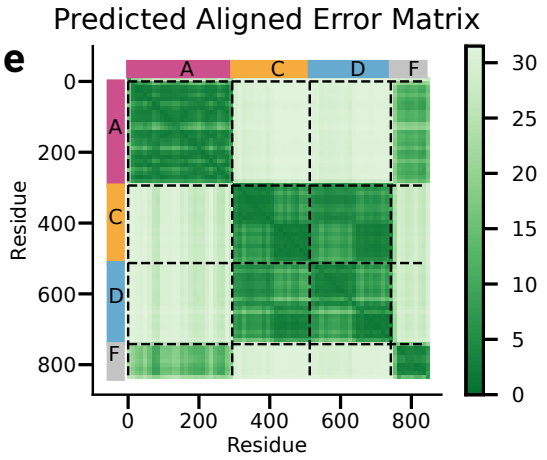


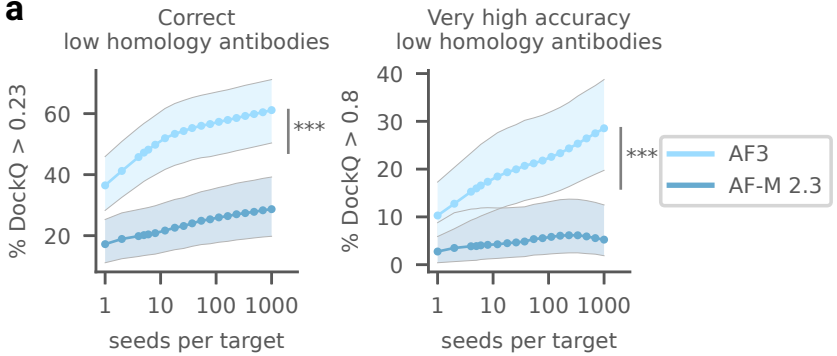


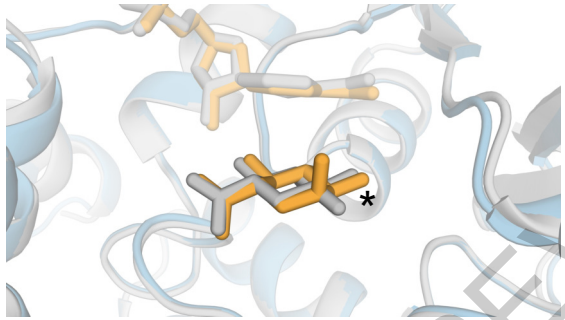


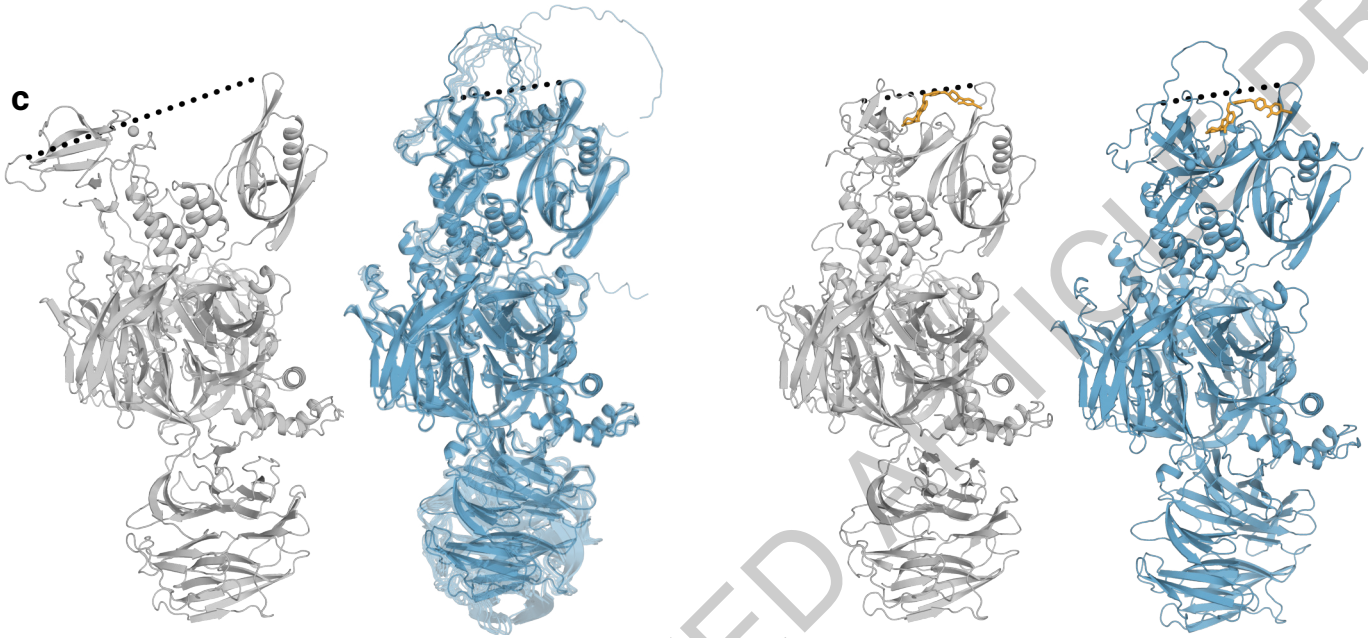


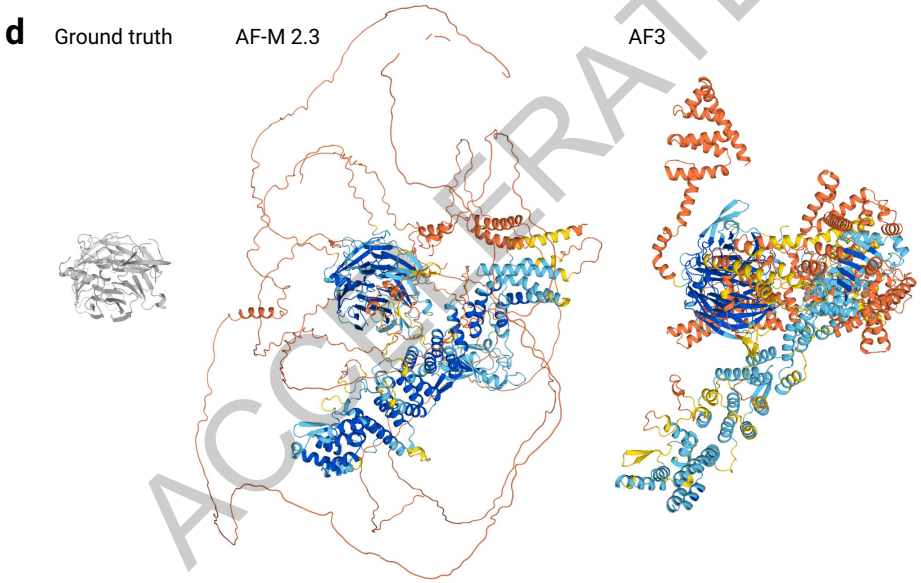


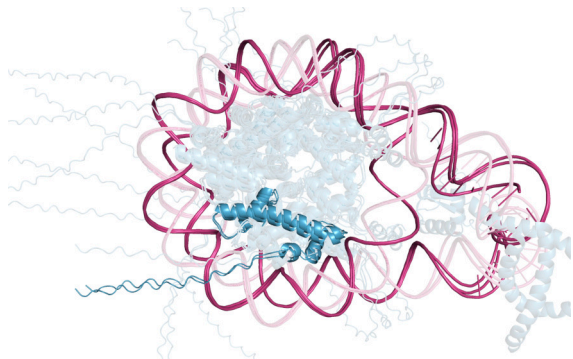


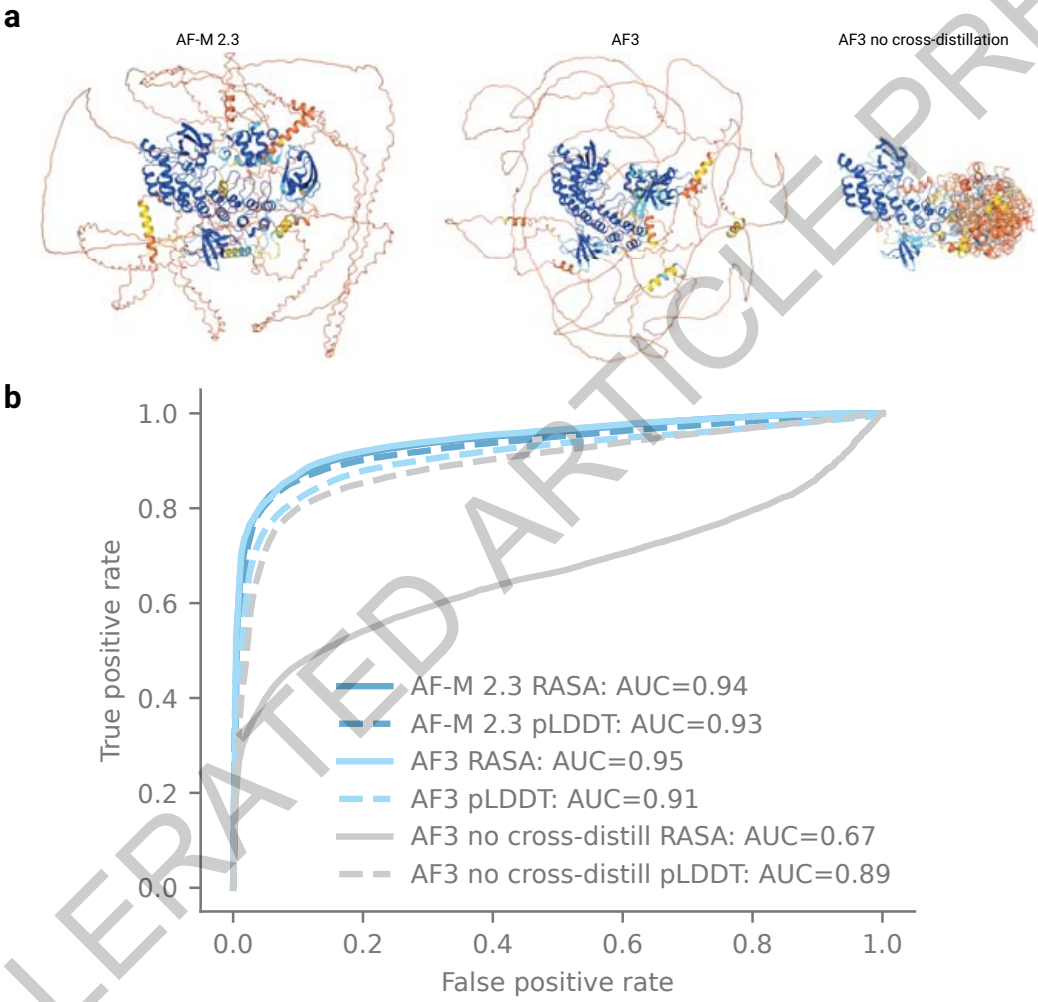




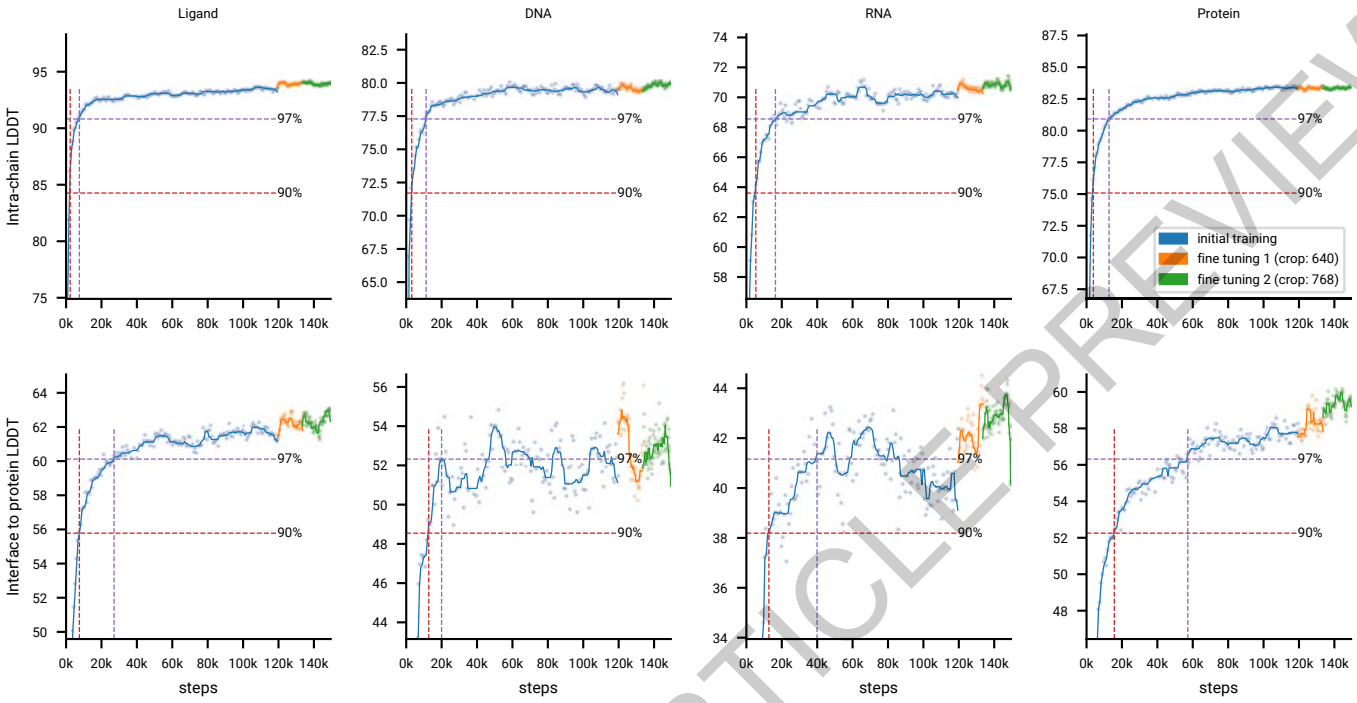




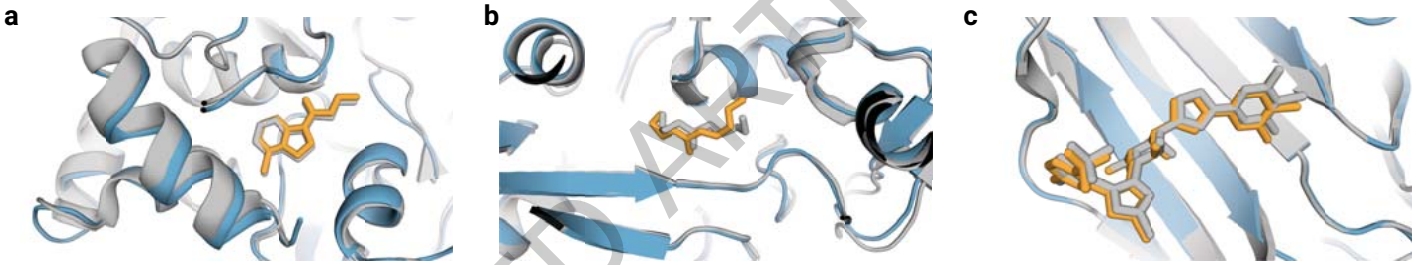




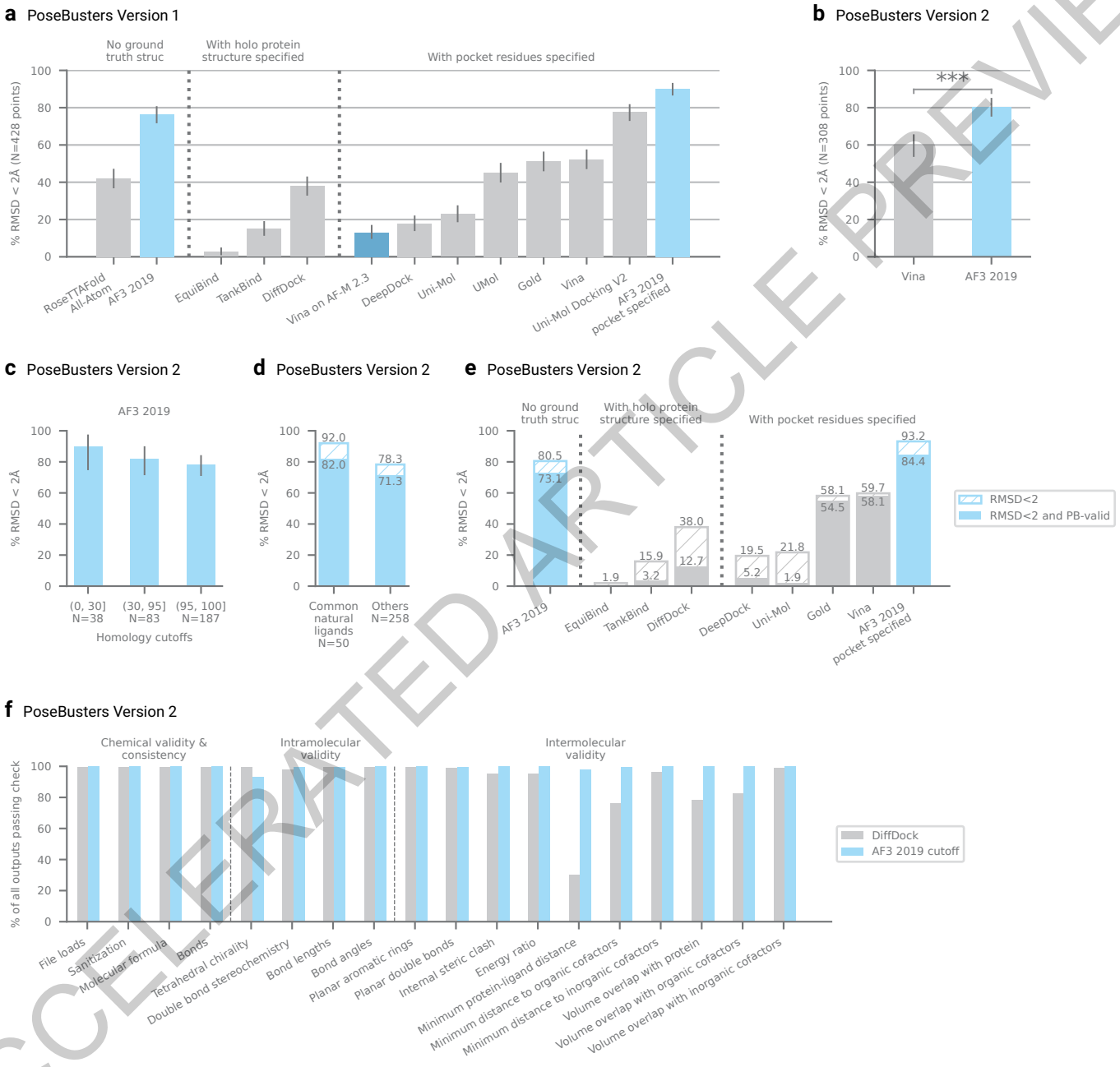
Extended Data Fig. 1



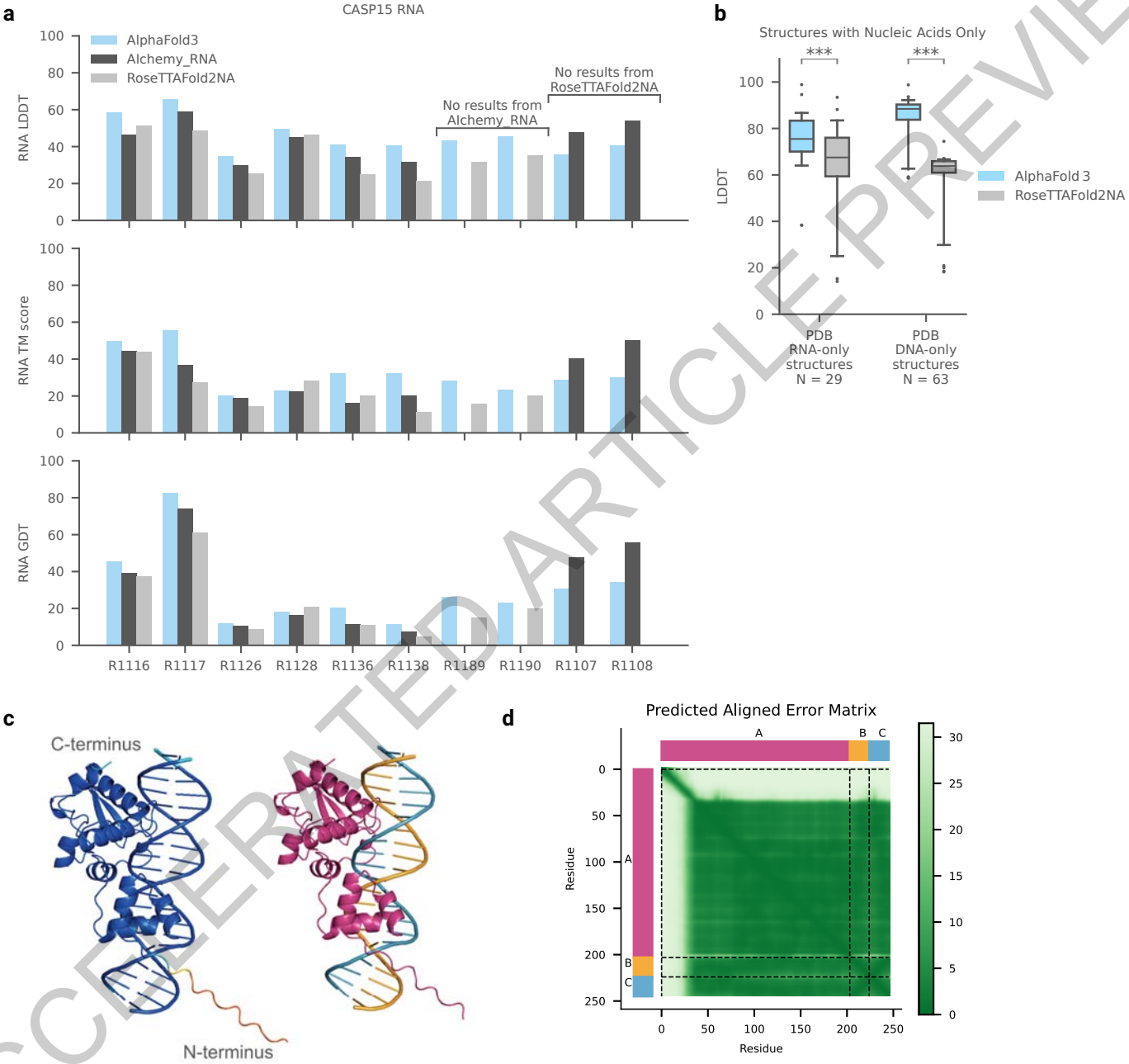
Extended Data Fig. 2



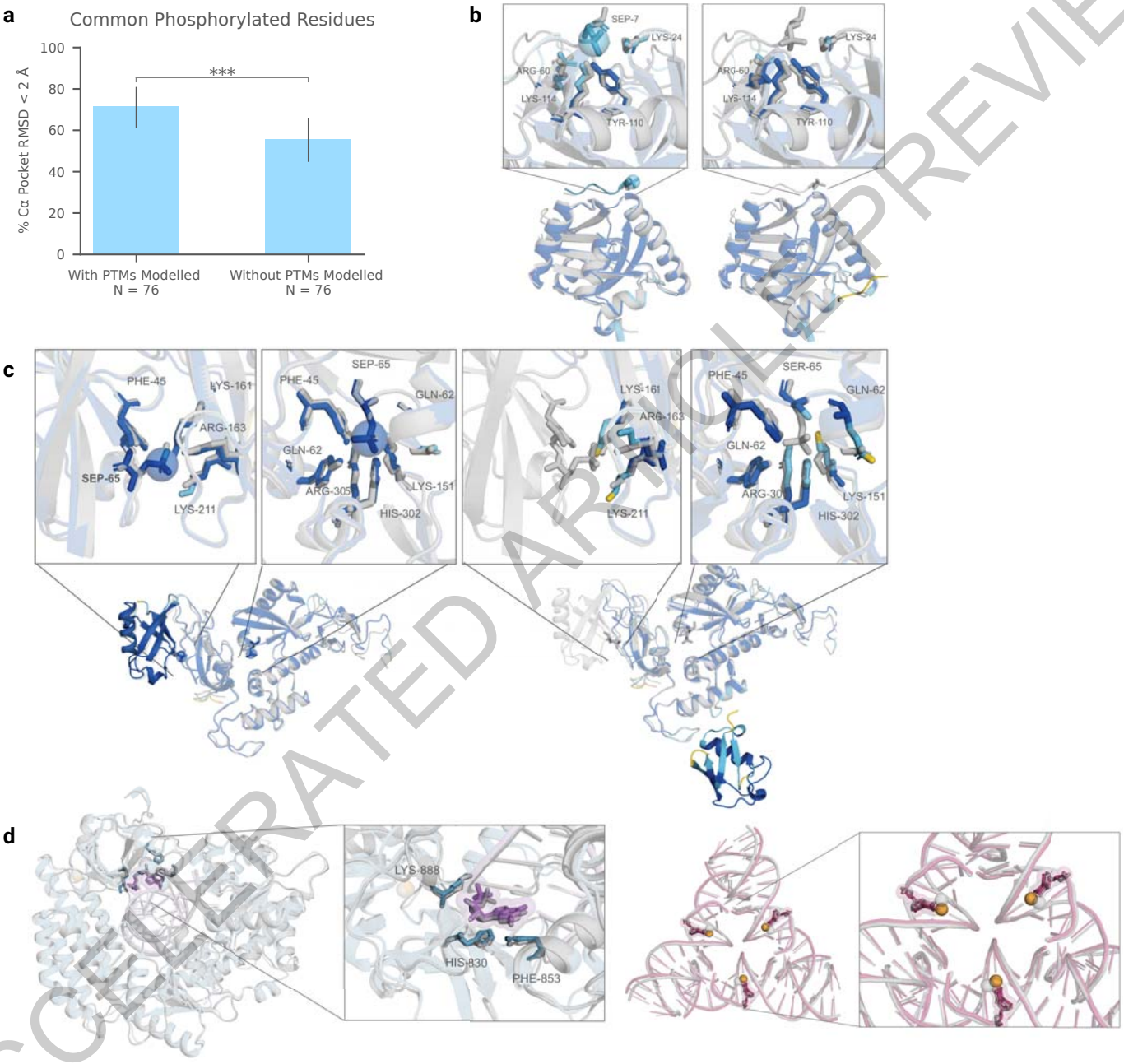
Extended Data Fig. 3



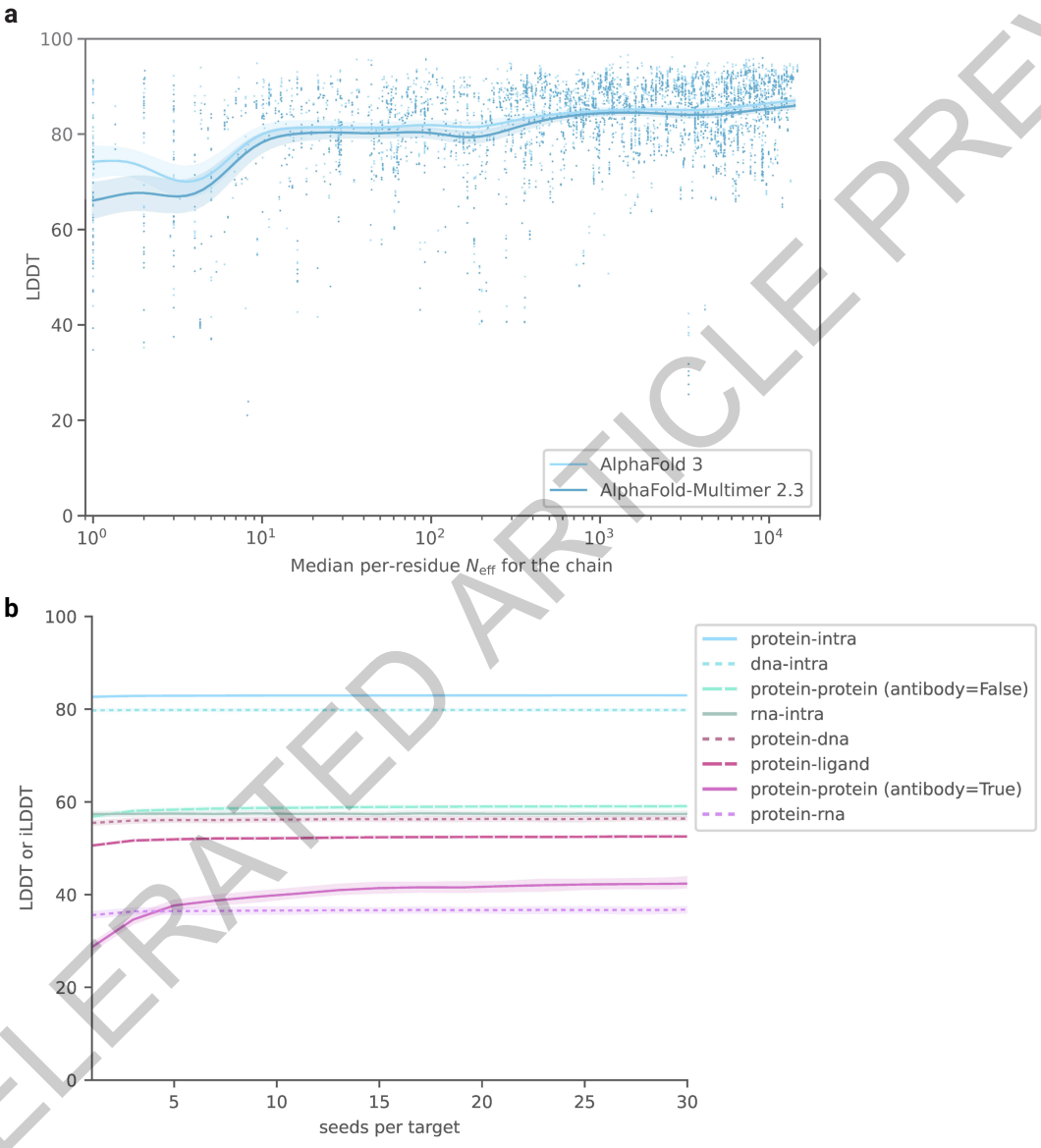
Extended Data Fig. 4



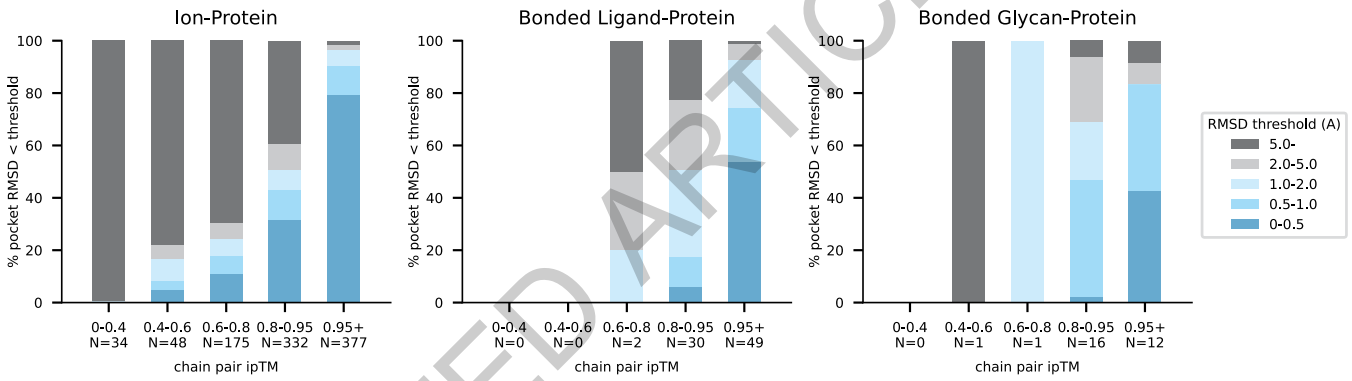
Extended Data Fig. 5



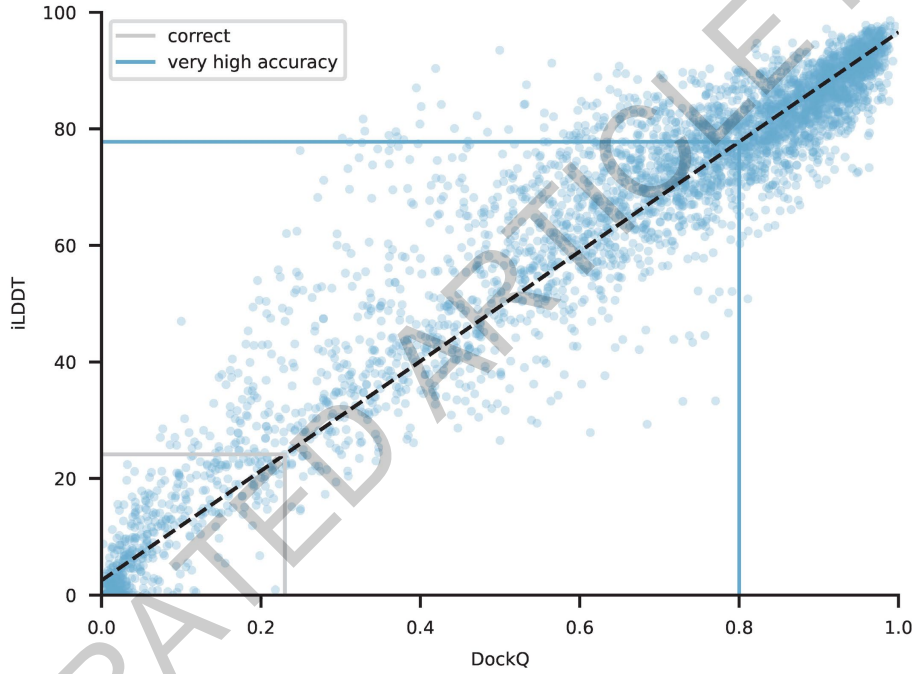
Extended Data Fig. 6



Extended Data Fig. 7



Extended Data Fig. 8



Extended Data Fig. 9

Task	Dataset	Metric	Notes	Method	N	Mean	95% CI
Ligands	PoseBusters V1	% RMSD < 2 Å	–	RoseTTAFold All-Atom	427	42.0	37.2 – 46.8
				AF3 (2019 cutoff)	428	**76.4**	72.1 – 80.3
			Holo protein struct. given	EquiBind	428	2.6	1.3 – 4.6
				TankBind	428	15.0	11.7 – 18.7
				DiffDock	428	37.9	33.2 – 42.6
			Pocket residues specified	Vina on AF-M 2.3	428	13.1	10.0 – 16.7
				DeepDock	428	17.8	14.3 – 21.7
				Uni-Mol	428	22.9	19.0 – 27.2
				UMol	428	45.0	40.3 – 49.9
				Gold	428	51.2	46.3 – 56.0
				Vina	428	52.3	47.5 – 57.2
				Uni-Mol Docking V2	428	77.6	73.3 – 81.4
				AF3 (2019 cutoff) pocket specified	428	**90.2**	87.0 – 92.8
	PoseBusters V2	% RMSD < 2 Å	–	AF3 (2019 cutoff)	308	**80.5**	75.6 – 84.8
			Holo protein struct. given	EquiBind	308	1.9	0.7 – 4.2
				TankBind	308	15.9	12.0 – 20.5
				DiffDock	308	38.0	32.5 – 43.7
			Pocket residues specified	Vina on AF-M 2.3	308	15.3	11.4 – 19.8
				DeepDock	308	19.5	15.2 – 24.4
				Uni-Mol	308	21.8	17.3 – 26.8
				Gold	308	58.1	52.4 – 63.7
				Vina	308	59.7	54.0 – 65.3
				AF3 (2019 cutoff) pocket specified	308	**93.2**	89.8 – 95.7
Nucleic Acids	Protein-RNA	iLDDT		RoseTTAFold2NA	25	19.0	15.6 – 23.2
				AF3	25	**39.4**	28.5 – 51.9
	Protein-dsDNA	iLDDT		RoseTTAFold2NA	38	28.3	20.7 – 37.5
				AF3	38	**64.8**	56.4 – 71.7
	CASP 15 RNA	RNA LDDT		RoseTTAFold2NA	8	35.5	28.3 – 43.8
				AF3	8	47.3	41.7 – 55.2
				AIchemy_RNA2 (has human input)	8	**54.5**	45.3 – 62.4
				RNApolis (has human input)	8	50.5	45.2 – 55.8
				Chen (has human input)	8	49.8	40.7 – 58.5
				Kiharalab	8	40.9	35.1 – 54.3
				UltraFold	8	37.8	32.5 – 45.0
Covalent Mod.	Bonded ligands	% RMSD < 2 Å		AF3	66	**78.5**	68.3 – 86.2
	Glycosylation	% RMSD < 2 Å	high-quality, single-residue	AF3	28	**72.1**	53.1 – 85.7
			all-quality, single-residue	AF3	167	**46.0**	40.0 – 52.1
			all-quality, multi-residue	AF3	131	**42.4**	35.4 – 49.3
	Modified residues	% RMSD < 2 Å		AF3	154	**59.9**	52.4 – 67.0
	Modified protein residues	% RMSD < 2 Å		AF3	40	**51.0**	36.0 – 65.6
	Modified DNA residues	% RMSD < 2 Å		AF3	91	**68.6**	59.0 – 76.9
	Modified RNA residues	% RMSD < 2 Å		AF3	23	**40.9**	23.4 – 59.9
Proteins	All Protein-Protein	% dockq > 0.23		AF-M 2.3	1064	67.5	64.7 – 70.1
				AF3	1064	**76.6**	74.0 – 78.9
	Protein-Antibody	% dockq > 0.23		AF-M 2.3	65	29.6	19.6 – 40.4
				AF3	65	**62.9**	51.4 – 73.5
	Monomers	LDDT		AF-M 2.3	338	85.5	84.7 – 86.1
				AF3	338	**86.9**	86.2 – 87.6

Extended Data Table 1

nature portfolio

Corresponding author(s): John Jumper (jumper@google.com)

Last updated by author(s): Mar 26, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All scientific datasets used to create training and evaluation inputs are freely available from public sources (see Data section below). No additional data was collected.
Data analysis	Data analysis used Python v3.11.7 (https://www.python.org/), NumPy v1.26.3 (https://github.com/numpy/numpy), SciPy v1.9.3 (https://www.scipy.org/), seaborn v0.12.2 (https://github.com/mwaskom/seaborn), Matplotlib v3.6.1 (https://github.com/matplotlib/matplotlib), pandas v2.0.3 (https://github.com/pandas-dev/pandas), statsmodels v0.12.2 (https://github.com/statsmodels/statsmodels), RDKit v4.3.0 (https://github.com/rdkit/rdkit), and Colab (https://research.google.com/colaboratory). TM-align v20190822 (https://zhanglab.dcmb.med.umich.edu/TM-align/) was used for computing TM-scores. Structure visualizations were created in Pymol v2.55.5 (https://github.com/schrodinger/pymol-open-source). PoseBusters scoring done with PoseBusters v0.2.7 (https://github.com/maabuu/posebusters). RoseTTAFold2NA benchmarking done with RoseTTAFold2NA v0.2 (https://github.com/uw-ipd/RoseTTAFold2NA).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All scientific datasets used to create training and evaluation inputs are freely available from public sources. Structures from the PDB were used for training and as templates (https://files.wwpdb.org/pub/pdb/data/assemblies/mmCIF/; for sequence clusters see https://cdn.rcsb.org/resources/sequence/clusters/clusters-by-entity-40.txt; for sequence data see https://files.wwpdb.org/pub/pdb/derived_data/).
Training used a version of the PDB downloaded 12 January 2023, while template search used a version downloaded 28 September 2022. We also used the Chemical Components Dictionary downloaded on 19 October 2023 (https://www.wwpdb.org/data/ccd).

We show experimental structures from the PDB with accession numbers 7PZB50,51, 7PNM52,53, 7TQL54,55, 7AU256,57, 7U8C58,59, 7URD60,61, 7WUX62,63, 7QIE64,65, 7T8266,67, 7CTM68,69, 8CVP43,70, 8D7U43,71, 7F6072,73, 8BTI74,75, 7KZ976,77, 7XFA78,79, 7PEU80,81, 7SDW82,83, 7TNZ84,85, 7R6R 86,87, 7USR88,89, and 7Z1K.90,91

We also used the following publicly available databases for training or evaluation. Detailed usage is described in Supplementary Methods 2.2{Genetic search} and Supplementary Methods 2.5.2{Distillation datasets}.
UniRef90 v.2020_01 (https://ftp.ebi.ac.uk/pub/databases/uniprot/previous_releases/release-2020_01/uniref/),
UniRef90 v.2020_03 (https://ftp.ebi.ac.uk/pub/databases/uniprot/previous_releases/release-2020_03/uniref/),
UniRef90 v.2022_05
https://ftp.ebi.ac.uk/pub/databases/uniprot/previous_releases/release-2022_05/uniref/),
Uniclust30 v.2018_08
(https://wwwuser.gwdg.de/~compbiol/uniclust/2018_08/),
Uniclust30 v.2021_03
(https://wwwuser.gwdg.de/~compbiol/uniclust/2021_03/),
MGnify clusters v.2018_12
(https://ftp.ebi.ac.uk/pub/databases/metagenomics/peptide_database/2018_12/),
MGnify clusters v.2022_05
(https://ftp.ebi.ac.uk/pub/databases/metagenomics/peptide_database/2022_05/),
BFD
(https://bfd.mmseqs.com),
RFam v.14.9
(https://ftp.ebi.ac.uk/pub/databases/Rfam/14.9/),
RNAcentral v.21.0
(https://ftp.ebi.ac.uk/pub/databases/RNAcentral/releases/21.0/),
Nucleotide Database (as of 23 February 2023)
(https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nt.gz),
JASPAR 2022
(https://jaspar.elixir.no/downloads/; see https://jaspar.elixir.no/profile-versions for version information),
SELEX protein sequences from Supplementary Tables92
(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8009048/),
SELEX protein sequences from Supplementary Tables93
(https://www.nature.com/articles/nature15518).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All available data were used for each benchmark. No subsampling was performed.
Data exclusions	PDB structures were excluded on the basis of size or homology as described in the text
Replication	Code and method details were carefully checked for completeness and replicability.
Randomization	The work constitutes in-silico analysis so all treatments (software packages) were applied to all relevant data for benchmarking.
Blinding	Test sets were held back from training but researchers were not blinded. Large test sizes (all recent PDB) were used instead to avoid overfitting. Fully blind tests would be impractical over the development of the project due to the small size of recent PDB and the need for large samples size on individual new prediction modalities.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*