Performance and structural coverage of the latest, in-development AlphaFold model

Google DeepMind AlphaFold Team[1] **and Isomorphic Labs Team**[2]
[1]DeepMind, London, UK, [2]Isomorphic Labs, London, UK

The introduction of AlphaFold 2 (Jumper et al., 2021) has spurred a revolution in modelling the structure of proteins and their interactions, enabling a huge range of applications in protein modelling and design (Kreitz et al., 2023; Lim et al., 2023; Mosalaganti et al., 2022). In this note, we report on our progress on a new iteration of AlphaFold modelling that greatly expands the range of applicability of the method and is capable of joint structure prediction of complexes including proteins, nucleic acids, small molecules, ions, and modified residues. The new AlphaFold model demonstrates greatly improved accuracy over previous specialist tools in the majority of cases: far greater accuracy on protein-ligand interactions than state of the art docking tools, much higher accuracy on protein-nucleic acid interactions than specialist predictors like RoseTTA2FoldNA (Baek et al., 2022), and significantly higher antibody-antigen prediction accuracy than AlphaFold-Multimer (Evans et al., 2021). In this results-only progress report, we show quantitative benchmarks and highlight a number of specific high-accuracy predictions on recently solved structures. We believe that these results ultimately point to the achievability of atomically-accurate structure prediction for the full range of biomolecular interactions across the PDB within an AlphaFold framework.

Contents

1. Introduction

AlphaFold 2 (Jumper et al., 2021) was a fundamental breakthrough in single chain protein structure prediction. Subsequently, AlphaFold-Multimer (Evans et al., 2021) expanded the system to apply to complexes, demonstrating the broader applicability of the method. A follow-up release (AlphaFold 2.3 (AlphaFold-Team, 2022)) in December 2022 expanded coverage to larger complexes. However, complexes containing non-protein elements were not addressed by any previous AlphaFold model. Examples of such non-protein elements include small molecule ligands, nucleic acids, and modified residues and are of great interest to biologists. These non-protein components are often essential to the function and regulation of the associated biological complex.

To address the complexity of full biological systems, we are developing a new generation of AlphaFold that can handle arbitrary interactions of proteins, nucleic acids, arbitrary small molecule ligands, and modified or non-canonical residues. Development is still in progress, but striking performance across a wide range of tasks has already been achieved using our latest snapshot. This model, which we term "AlphaFold-latest" in this manuscript, obtains state of the art accuracy across

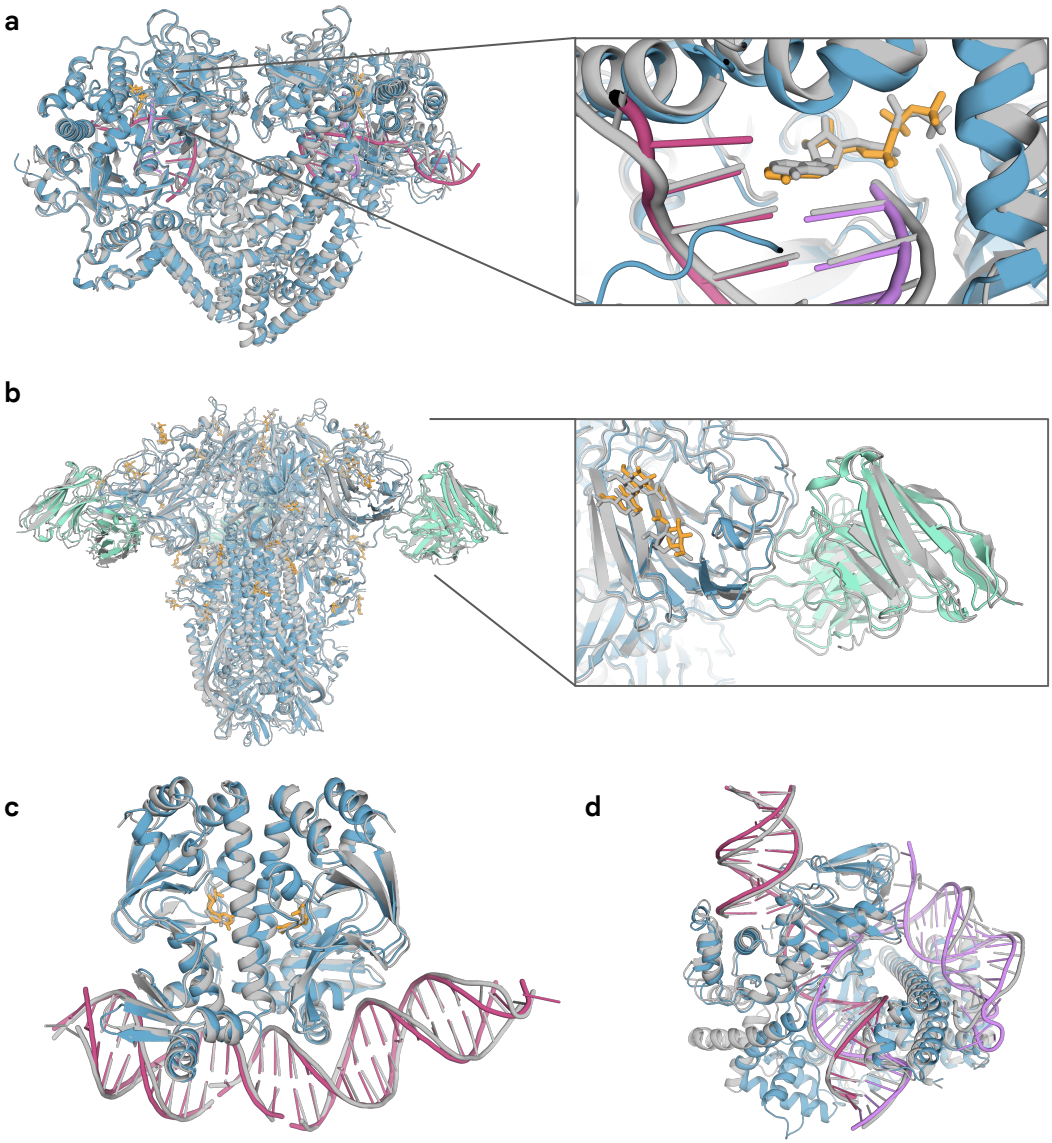

Figure 1 | Structure predictions from AlphaFold-latest. Predicted protein chains are shown in blue (predicted antibody in green), predicted ligands and glycans in orange, predicted DNA in pink, predicted RNA in purple, and ground truth in grey. **a**, two-barrel RNA polymerase QDE-1 bound to RNA primers and non-hydrolyzable nucleotide, AMPNPP (PDB ID 7Y7S, full complex LDDT: 87.9). **b**, human coronavirus OC43 spike protein, 4665 residues, heavily glycosylated and bound by neutralising antibodies (PDB ID 7PNM, full complex LDDT: 83.3). **c**, bacterial CRP/FNR family transcriptional regulator protein bound to DNA and cGMP (PDB ID 7PZB, full complex LDDT: 82.4). **d**, CRISPR-Casλ protein in complex with CRISPR-RNA and DNA (PDB ID 8DC2, full complex LDDT: 76.7).

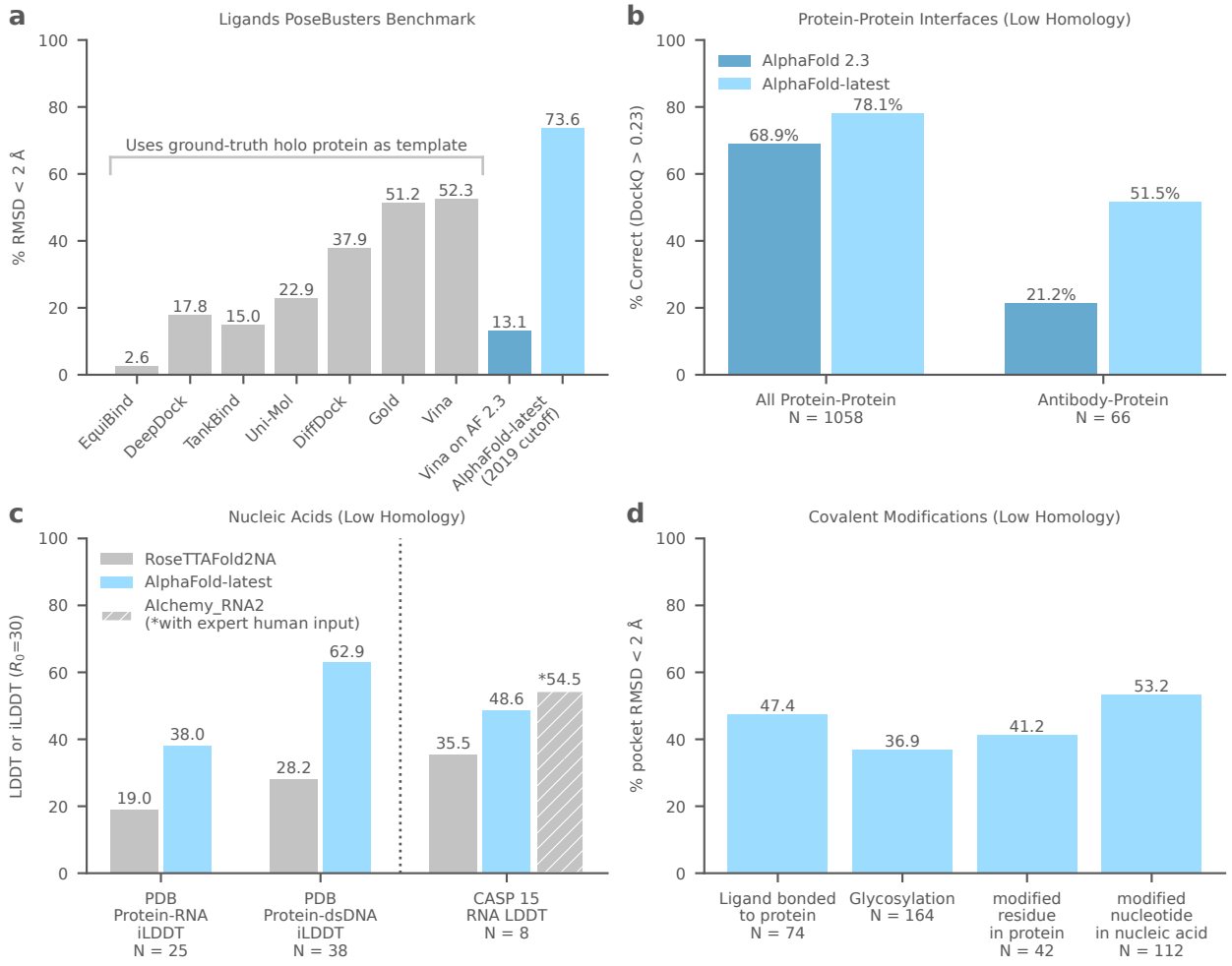

Figure 2 | Summary of AlphaFold-latest capabilities and performance. **a**, Ligand docking performance on PoseBusters benchmark set. N=428 targets. **b**, Protein-protein interaction accuracy **c**, Nucleic acid interaction and RNA accuracy. Nucleic acid LDDT is computed with an inclusion radius (R_0) of 30 Å. **d**, Accuracy on various covalent modifications.

nearly all categories of interactions. A range of selected predictions made by AlphaFold-latest are shown in Figure 1.

Figure 2 demonstrates performance in four categories:

(a) AlphaFold-latest outperforms classical systems like AutoDock Vina (Eberhardt et al., 2021; Trott and Olson, 2009) on the PoseBusters benchmark (Buttenschoen et al., 2023) for ligand docking despite baselines using ground truth bound protein structures as inputs while AlphaFold-latest starts from the protein sequences and ligand identities only. See Figure 6 for a range of example ligand predictions.

(b) It improves upon AlphaFold 2.3 for protein-protein structure prediction, especially in certain categories such as antibody binding structures.

(c) On protein-nucleic acid interfaces AlphaFold-latest outperforms competing systems (Baek et al., 2022), while for RNA structure prediction it outperforms automated methods but is slightly below the top CASP15 entrant which uses manual expert intervention (AIchemy_RNA2) (Chen et al., 2022; Xiong et al., 2021).

(d) Finally, AlphaFold-latest is able to predict the structure of further entities like bonded ligands,

glycosylation, and modified residues or nucleotides.

In the following sections we will present results for each molecule type, analyse our confidence measures and detail our evaluation procedure. This work is on-going, but demonstrates the progress being made towards the goal of a structure prediction system that makes functionally useful predictions for all biologically-relevant molecules.

2. Model Inputs and Outputs

AlphaFold-latest takes as input a description of the biological assembly, with sequences for polymers and SMILES for ligands, and optionally the sequence location of covalently bonded ligands, and outputs a prediction for the 3D position of each heavy atom. Water and hydrogens are excluded. All experimental structures used for training the model were from PDB with release dates up to 2021-09-30. Templates were filtered to only those released prior to 2021-09-30.

Inputs are "tokenized" to get model inputs, with one token per standard polymer residue and one token per heavy atom for ligands and nonstandard polymer residues. The number of tokens is the primary driver of compute time and limits of prediction sizes on different hardware. We evaluate system performance on complexes up to 5,120 tokens for computational ease, but the system is capable of running larger complexes on accelerators with large amounts of memory.

Each output structure comes with per-atom, per-token-pair, and aggregated structure-level confidence measures. In addition, each entity within the structure and each interface between entities within the structure has an associated confidence measure.

3. Results

All results are generated by running the model with 5 different random seeds. Confidence measures are used to rank predictions for the overall structure or for each chain or interface.

We constructed a large PDB evaluation set consisting of structures past our training date cutoff (see subsection A.1 for details). A subset of this set was determined to be low homology (see subsection A.2) to our training data and performance related results focus on that subset. Results are aggregated over clusters of similar entities (see subsection A.3), and cluster count in each result are shown in our figures (see subsection A.5).

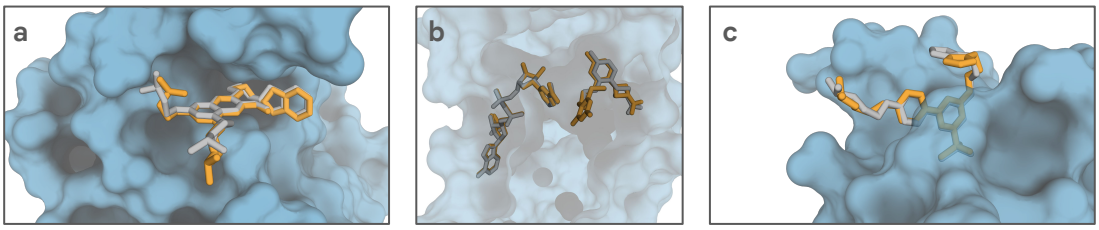

Figure 3 | Three AlphaFold-latest examples from the PoseBusters benchmark set where docking programs Vina and Gold fail to achieve accurate predictions. Surface representation of the predicted protein structure shown in blue, predicted ligand pose shown as sticks in orange, ground truth ligand pose shown as sticks in grey. **a**, PDB ID 7OCB: best docking RMSD = 4.6 Å, AlphaFold-latest RMSD = 0.96 Å. **b**, PDB ID 5SD5: best docking RMSD = 4.5 Å, AlphaFold-latest RMSD = 0.92 Å. **c**, PDB ID 7BLA: best docking RMSD = 6.3 Å, AlphaFold-latest RMSD = 2.0 Å.

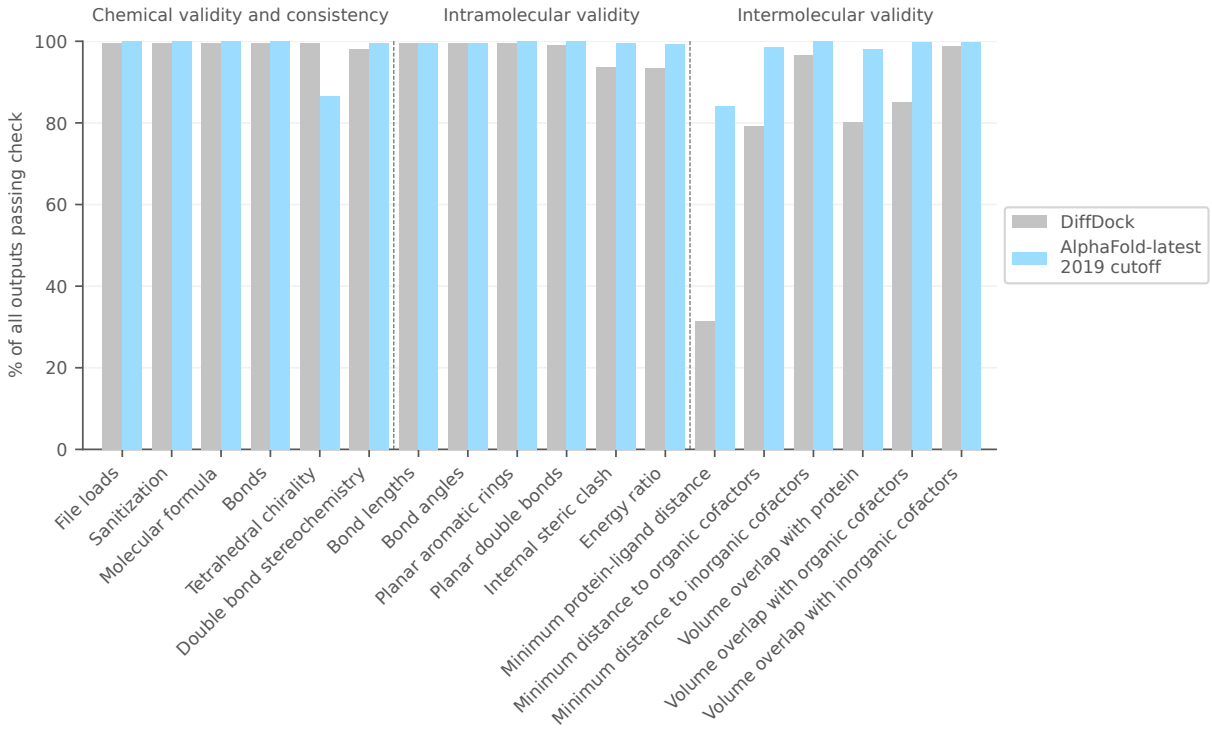

Figure 4 | Percentage of predictions passing quality checks from the PoseBusters tool, compared to an ML baseline.

In addition to results on this PDB evaluation set, we report results on PoseBusters for ligands and on publicly-available CASP15 RNA targets.

3.1. Ligands

We evaluated ligand accuracy on two datasets. First we looked at the PoseBusters benchmark set (Buttenschoen et al., 2023), a curated collection of 428 liganded protein structures from the PDB. Since this benchmark includes examples released before our usual training cut-off date, we generated predictions with a separately trained AlphaFold-latest model with an earlier cutoff date of 2019-09-30. Predictions were generated for each target's asymmetric unit based on the contents of the mmCIF file deposited in PDB and did not use any templates released after 2019-09-30. In a small number of cases the PDB file contained chains clashing with the evaluated ligand (both being present with some occupancy), in which case the clashing chains were removed. Only one target had too many tokens to run in full (the photosystem 8F4J), in which case we kept chains within 20 Å of the evaluated ligand. Evaluation was performed by running our usual chain assignment procedure, determining the protein chain with the most atoms within 10 Å of the ligand, then performing RMSD alignment of our prediction to the PDB mmCIF file on the backbone atoms of that protein chain. Our predicted ligand pose was then extracted and scored with the publicly available PoseBusters command line tool.

We report the success rate (RMSD under 2 Å) in Figure 2a. For baselines we report the values from the PoseBusters paper SI. Using the AlphaFold-latest model led to a 73.6% success rate. For comparison, the docking baseline Vina has a 52.3% success rate, and the machine learning baseline DiffDock has a 37.9% success rate. Note that both Vina and DiffDock are provided with the exact protein coordinates of the docked structure in this baseline, instead of a more realistic benchmark

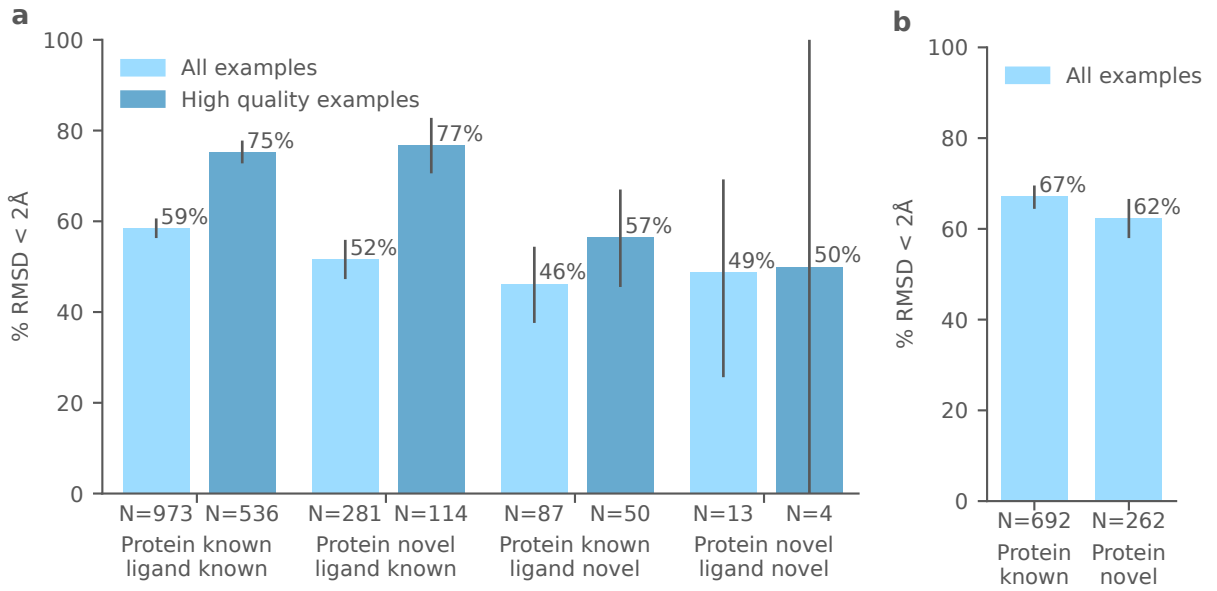

Figure 5 | AlphaFold-latest performance on recent PDB ligands. To produce each bar, the evaluation set was first filtered based on homology and (optionally) quality. A protein is considered novel based on a 40% template identity cut off, and a ligand based on a 0.5 Tanimoto similarity coefficient cut off relative to train (using 2048 bit RDKit fingerprints). Examples were assigned to a protein similarity cluster based on their main protein chain; the reported success rate is the average over these clusters. Error bars represent 95% bootstrap confidence intervals. **a**, general ligands. **b**, ions.

using homology models or protein coordinates from complexes with different ligands; Vina also receives a specification of the binding region in the form of a 25 Å wide box (local docking). In contrast, AlphaFold-latest requires only protein sequence and ligand identity information without any pocket information. Figure 3 shows three examples where AlphaFold-latest achieves accurate predictions but docking tools Vina and Gold do not. Finally, we confirmed the previously observed (Karelina et al., 2023; Scardino et al., 2023) poor performance of attempting to use classical docking tools to dock into AlphaFold 2.3 protein structures; instead our work demonstrates that AlphaFold-latest can make dramatically better predictions for those bound structures when jointly predicting the protein and ligand positions.

We also looked at the various quality checks returned by the PoseBusters tool (Figure 4). Note that AlphaFold-latest performs structure prediction, not rigid docking, and so it may generate a prediction with local changes in the surrounding context to accommodate the ligand. As such, it is more appropriate to evaluate the intermolecular plausibility of AlphaFold-latest predictions by checking for clashes between the predicted ligand and its predicted context, rather than between the predicted ligand and the ground truth protein. This is the approach taken in Figure 4 for the intermolecular validity checks only.

The second dataset we considered was our usual recent PDB evaluation set. For this set, ligands were filtered to remove anything nonstandard: glycan components (Table 2), covalently bonded ligands, ions (Table 3), and anything represented by multiple three letter codes in the PDB. To enrich for biologically and therapeutically relevant ligands, a further exclusion list was applied (Table 1).

A pocket aligned RMSD score was computed for each ligand (see subsection A.4). Each ligand was further assigned a main protein chain (the one with which it has the largest interface), to be used in clustering and homology filtering. For this investigation, filtering by protein novelty and ligand novelty are considered separately.

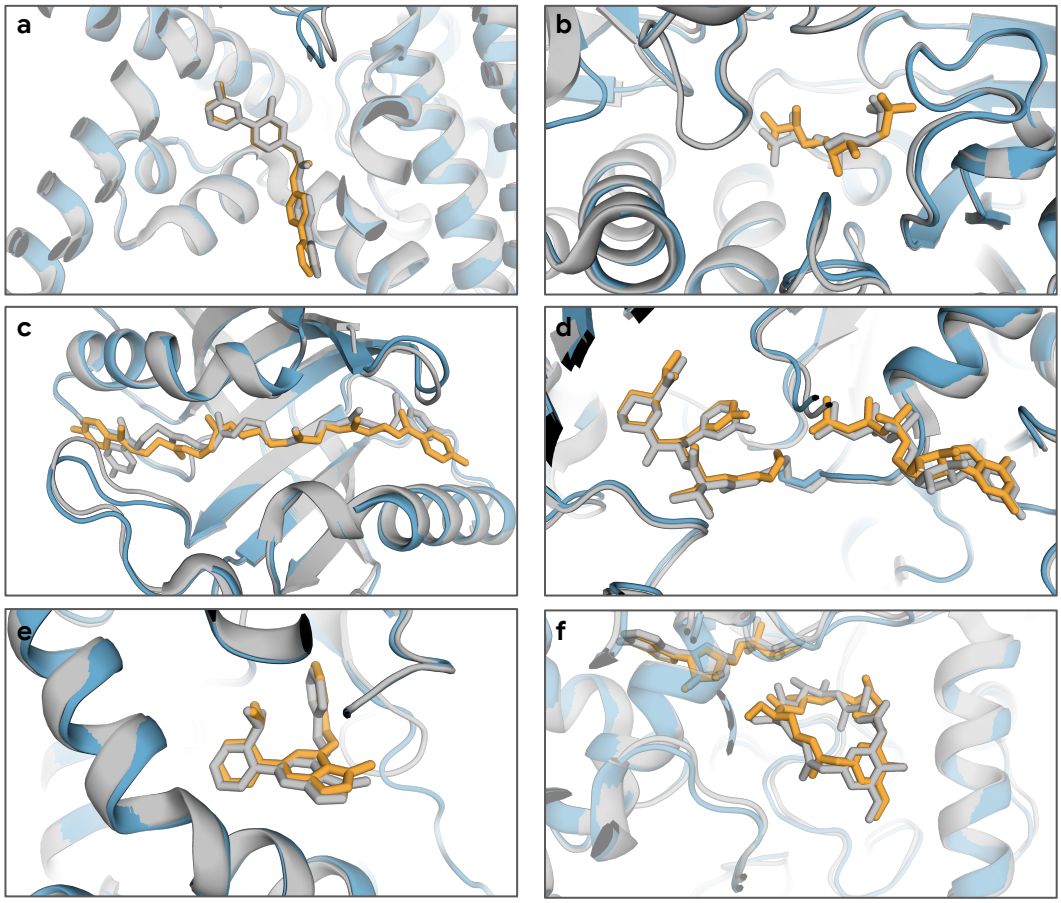

Figure 6 | Case studies highlighting the performance of AlphaFold-latest on recent therapeutically relevant structures. Structures of the predicted protein chains are shown in blue, predicted ligand poses are shown as sticks in orange, and ground truth structures are in grey. **a**, LGK974, a clinical stage inhibitor, bound to PORCN in complex with the WNT3A peptide (PDB ID 7URD, ligand RMSD 0.39 Å). **b**, (5S,6S)-O7-sulfo DADH bound to the AziU3/U2 complex with a novel fold (PDB ID 7WUX, ligand RMSD 1.19 Å) . **c**, Closthioamide bound to CtaZ (PDB ID 7ZHD, ligand RMSD 2.22 Å). **d**, Sanglifehrin A analog covalently bound to KRAS G12C and immunophilin CYPA, in a tri-complex with cofactor (PDB ID 8G9Q, ligand RMSD 0.80 Å, location of covalent bond not specified a priori). **e**, Analog of NIH-12848 bound to an allosteric site of PI5P4Kγ (PDB ID 7QIE, ligand RMSD 0.85 Å). **f**, GdmN in complex with the macrocycle 20-O-methyl-19-chloroproansamitocin and cofactor (PDB ID 7VZN, ligand RMSD 1.02 Å).

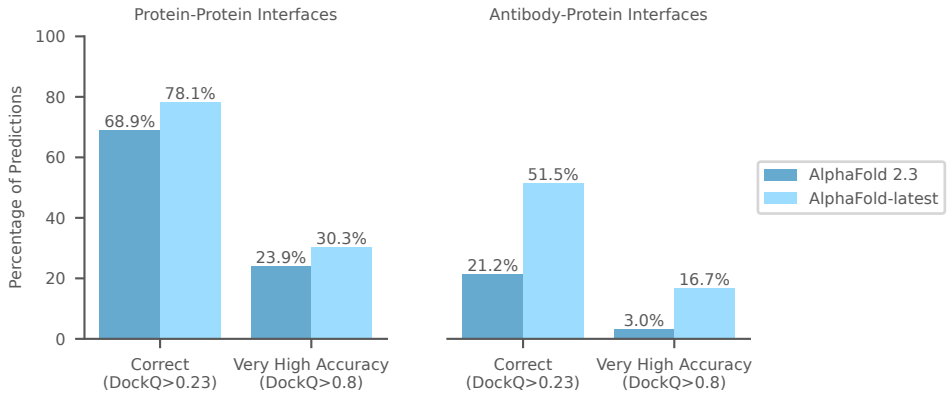

Figure 7 | Protein-protein prediction performance on the low homology eval set. The set contains 1,058 low homology interface clusters, 66 of which are antibody-protein interfaces; reported figures are mean within cluster, averaged across clusters.

Figure 5a shows the success rate (RMSD < 2 Å) of the AlphaFold-latest model on our recent PDB set. When all evaluation examples are considered, the success rate varies between 59% and 46% depending on the degree of novelty. However, this set does not take into account the quality of the underlying experimental data. Therefore we also show the effect of applying a simple quality filter: including just ligands with a ranking_model_fit > 0.5 according to the RCSB structure validation report (that is, X-ray structures with a model quality above the median) . When this filter is applied, the success rate increases notably to reach a range between 77% and 50%, similar to the results on the quality-filtered PoseBusters set. This underlines the impact that different data-cleaning and filtering procedures can have when evaluating ligand prediction systems. Figure 5b shows the same analysis for ions (see Table 3 for the full list of included CCD codes (codes from the Chemical Component Dictionary, an external reference file referring to all the residues and molecules found in the PDB). As all ion types in our eval set also exist in the training set, we only present the two results for known and novel proteins. The success rates of 67% and 62% are higher than for general ligands, perhaps because there exist frequent, well-preserved binding motifs, like zinc fingers.

In Figure 6 we show a selection of qualitative examples highlighting the ability of the model to generalise to a number of therapeutically relevant proteins and modalities. As mentioned above, AlphaFold-latest used only the protein sequence and ligand identity information, without ground truth protein coordinates or any pocket information. The examples cover molecular glues and covalent inhibitors (KRAS/CYPA), allosteric sites (PI5P4Kγ), integral membrane proteins (PORCN), macrocycles (GdmN), as well as proteins with unique folds and previously uncharacterized binding sites (AziU3/U2). The case studies span both current human biology targets (PORCN/WNT3A) and potential anti-infective targets (CtaZ), while the diversity of the set highlights the model's potential to structurally characterise some of the most challenging drug targets of today. In selecting these examples, we considered novelty in terms of the similarity of the protein to the training set (<40% sequence identity to the training set - GdmN, CtaZ, PORCN, AziU3/U2), the similarity of the ligand to the training set (<0.5 Tanimoto similarity - PI5P4Kγ), or the novelty of the protein-ligand pair (<0.5 Tanimoto similarity to ligands assigned to this protein - KRAS/CYPA).

3.2. Proteins

In Figure 7 and Figure 2b, we compare the performance of AlphaFold-latest with AlphaFold 2.3 on the low homology recent PDB evaluation set. For comparison against AlphaFold 2.3, the set

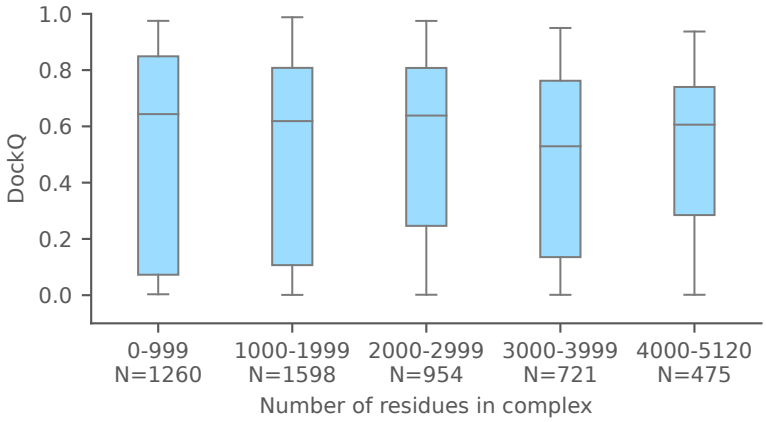

Figure 8 | Interface prediction accuracy - breakdown by complex size. Median DockQ for protein-protein interfaces in the recent PDB evaluation set is shown; this analysis is done without clustering the interfaces as the same interface can appear in different sized complexes. The counts represent the number of scored interfaces in each bin.

was restricted to complexes with at most 2,560 protein residues, at most 20 protein chains and greater than 3 residues in a chain. A further 62 interfaces (less than 2% of the total) were filtered out due to technical complications when running AlphaFold 2.3. All comparisons between the two systems only use per-complex confidence for ranking. On the low homology recent PDB evaluation set, AlphaFold-latest outperforms AlphaFold 2.3 by correctly (DockQ>0.23) predicting the interface 78.1% (+9.2) of the time and producing very high accuracy (DockQ>0.8) interface predictions in 30.3% (+6.4) of cases. For antibody-antigen interface prediction, the improvement is even greater, with AlphaFold-latest correctly predicting the interface in 51.5% (+30.3) of cases, with very high accuracy predictions for 16.7% (+13.7) of examples.

In Figure 8 we show that accuracy of interface predictions remains high for large complexes. Prediction accuracy for monomeric proteins remains high for AlphaFold-latest, with a median LDDT over low homology clusters of 88.4, compared to 87.2 for AlphaFold 2.3.

3.3. Nucleic Acids

AlphaFold-latest is capable of predicting nucleic acid (DNA and/or RNA) structures, either alone or in complex with proteins and/or ligands.

In Figure 9a, we present a detailed breakdown of nucleic acid performance over low homology clusters for protein-DNA interfaces, RNA chains, and protein-RNA interfaces. Nucleic acid accuracy was assessed via LDDT for intra chains and interface LDDT (iLDDT) for interfaces (see subsection A.4 below). Each complex type was further defined as small (up to 1000 total amino acids and nucleotides) or large (greater than 1000 residues). For protein-DNA interfaces, our mean performance across clusters is 61.0 iLDDT for small complexes and 31.2 for large complexes. RNA chain clusters have a mean LDDT of 65.0 for small complexes and 58.9 for large complexes. Protein-RNA interface clusters have a mean iLDDT of 39.7 for small complexes and 26.5 for large complexes.

In Figure 9b, we compare AlphaFold-latest to a recent deep learning based system for general protein-nucleic acid prediction, RoseTTAFold2NA (RF2NA) (Baek et al., 2022). For side-by-side comparison, we present average interface scores per target instead of per cluster as in Figure 9a. We evaluate protein-DNA interfaces split into paired and unpaired sets. Paired DNA is defined as complexes which have perfectly base-pair complementary chains. The remaining systems with DNA

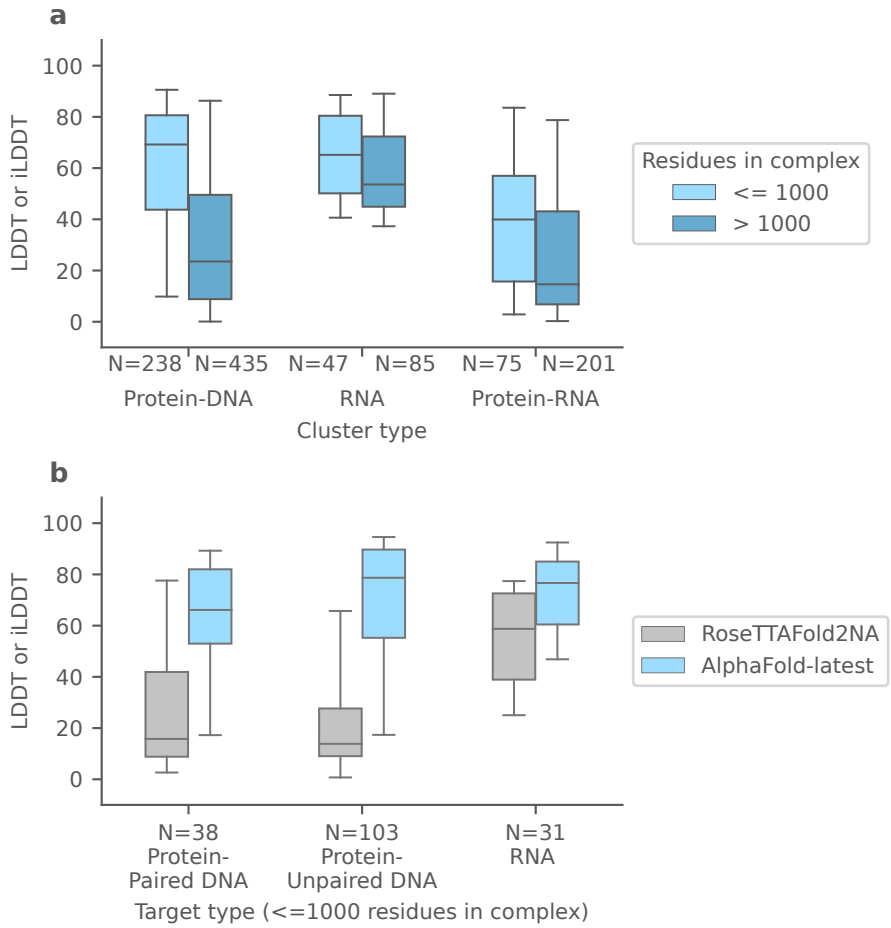

Figure 9 | Nucleic acid complex performance. **a**, Performance of AlphaFold-latest at predicting low homology protein-DNA interfaces, RNA chains, and protein-RNA interfaces from our PDB test set (averages per interface cluster, split by as many as or more than 1000 total complex residues - amino acids and nucleotides). N refers to the number of interface clusters. iLDDT used for Protein-DNA and Protein-RNA; LDDT for RNA chains. **b**, Performance of AlphaFold-latest vs RoseTTAFold2NA on low homology nucleic acid targets with up to 1000 residues from our PDB test set. N refers to the number of targets, used instead of interface clusters to separate targets with and without paired DNA. iLDDT used for Protein-DNA; LDDT for RNA chains. Box, center line, and whiskers boundaries are at (25%, 75%) intervals, median, and (5%, 95%) intervals.

are defined as unpaired (have single-stranded residues or chains). RNA performance is reported only on RNA chains as most low homology complexes were too large to run on RF2NA (e.g. only 25 targets were available to make Figure 2c). AlphaFold-latest exceeds RF2NA performance, with a 34.7 mean iLDDT difference on paired protein-DNA interfaces, a 46.8 mean iLDDT difference on unpaired protein-DNA interfaces, and a 16.9 mean LDDT difference on RNA. The large difference in unpaired DNA complex performance is expected as RF2NA was only trained and evaluated on DNA duplexes (at least 10 cross-chain hydrogen bonds).

All protein-nucleic acid binding interfaces are assumed to be residue-specific and therefore sequence-specific. For this analysis, non–specific predicted structures receive low iLDDT scores though they may be biologically plausible (in particular for large complexes such as nucleosomes).

In Figure 2c above, we evaluate AlphaFold-latest performance on CASP15 (Kryshtafovych et al., 2023) RNA targets that are currently publicly available. AlphaFold-latest outperforms other automated

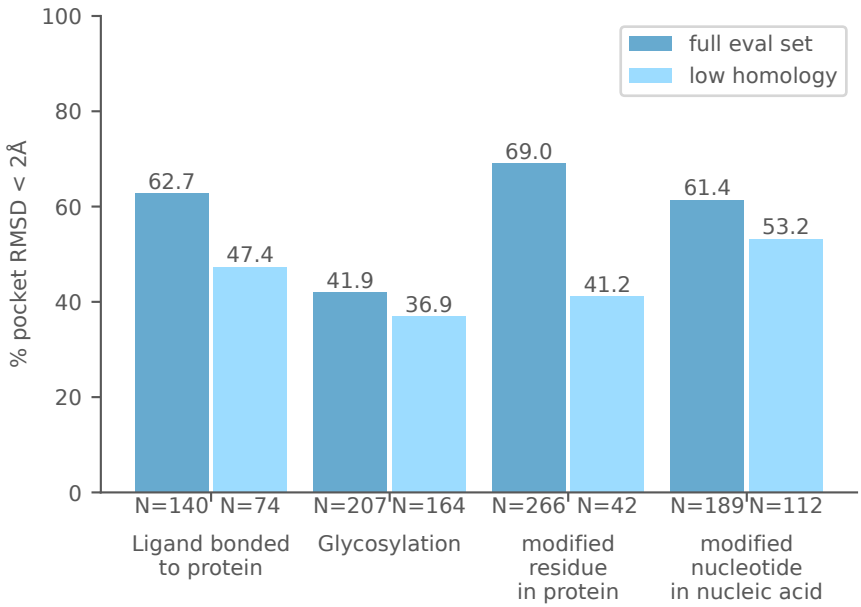

Figure 10 | Performance of AlphaFold-latest on modified structures. The values show the fraction of successful predictions (pocket aligned RMSD < 2 Å, see text). N denotes the number of interface clusters.

methods but performs slightly worse than the best entrants with manual expert intervention.

3.4. Covalent Modifications

AlphaFold-latest can also predict structures containing covalent modifications (see example in Figure 1b). Covalent modifications are specified in the input to AlphaFold in the same way as they are represented in the PDB, i.e. they can either be defined as residues with non-standard CCD codes (e.g. “SEP” for phosphoserine) or by additional entries in the bonds table, e.g. a covalent bond between a ligand and a residue.

We measure the prediction quality of the modifications in the same way as ligands using the pocket aligned RMSD (see subsection A.4). The atoms that contribute to the RMSD are either all heavy atoms of the ligand or glycan (in case of bonded ligands or glycosylation) or all heavy atoms of a modified residue (in case of modified residues). Figure 10 shows the performance of AlphaFold-latest on our recent PDB evaluation set (see subsection A.1) for “bonded ligand”, “glycosylation” and “modified residues”. The “bonded ligands” set contains all ligands that are bonded to a protein chain, and that are not in the glycosylation set and not in Table 1. The “glycosylation” set contains all glycan chains that are bonded to a protein chain. A glycan chain must only contain subunits defined in Table 2. The “modified residues” set are residues (protein, RNA, DNA) that are not one of standard residues nor “unknown” (see Table 4). As mentioned in subsection A.1 ligands or glycans bonded to homomeric subcomplexes but lacking the corresponding homomeric symmetry were excluded.

3.5. Confidences

As with AlphaFold 2, AlphaFold-latest provides a confidence of its prediction. Here we show calibration of AlphaFold-latest predicted error – a model-internal error prediction that can be thought of as an inverse confidence – with overall structure accuracy. Our analysis is performed on the recent PDB evaluation set, with no homology filtering and including peptides. All statistics are cluster-weighted (as described in subsection A.5 below).

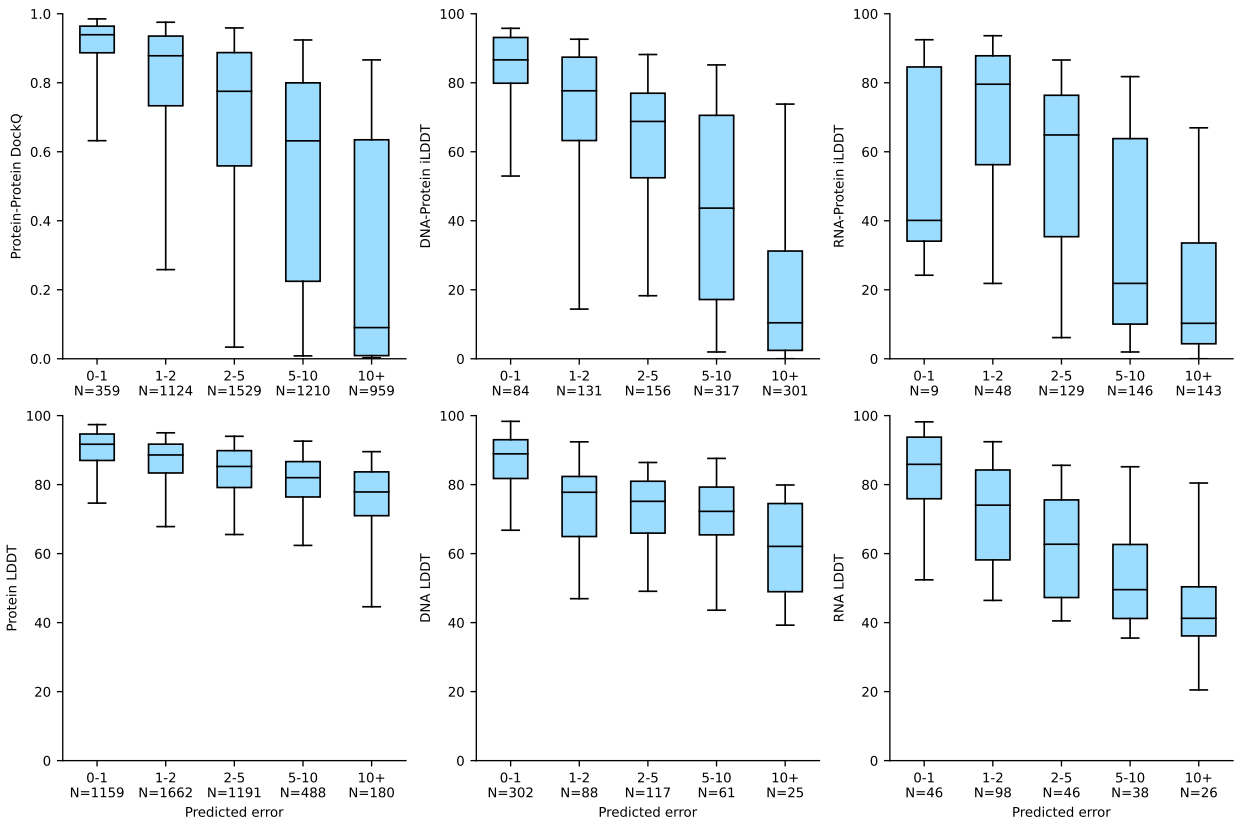

Figure 11 | Confidence vs. accuracy calibration across polymers. Accuracy of polymer-polymer interfaces (top row) and polymer chains (bottom row) as a function of predicted error. Box, centerline, and whiskers boundaries are at (25%, 75%) invtervals, median, and (5%, 95%) intervals.

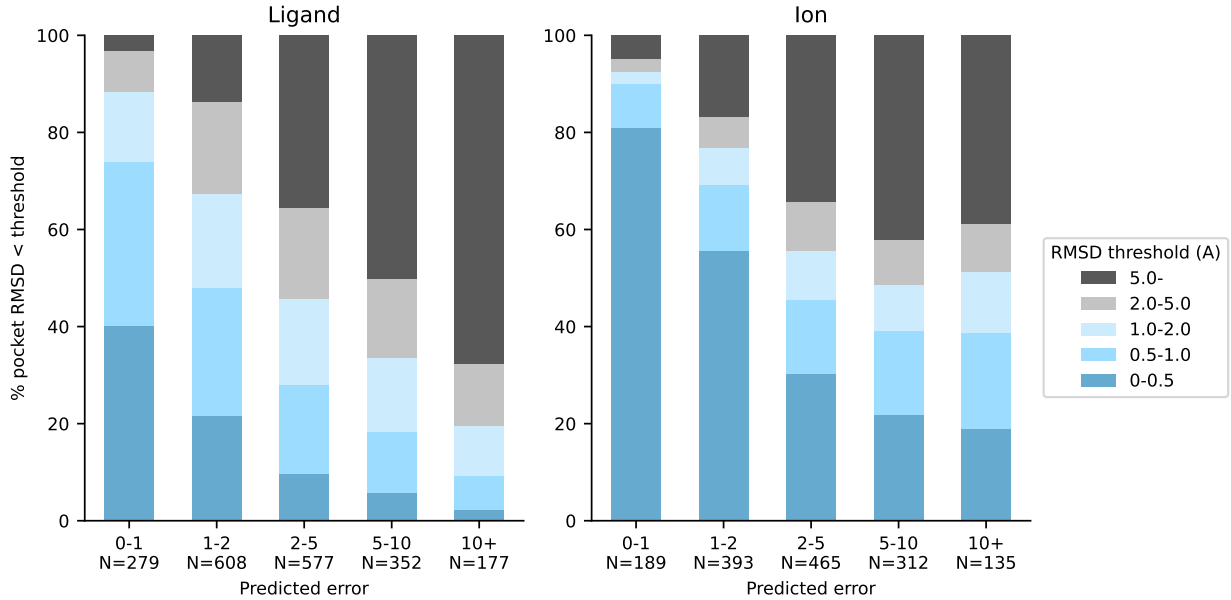

Figure 12 | Confidence vs. accuracy calibration for ligand-protein (left) and ion-protein (right) interfaces: Percentage of interface clusters under various pocket-aligned RMSD thresholds, as a function of the predicted error. Note that the predicted error is not a prediction of pocket aligned RMSD, although values are in similar ranges. N values report the number of clusters in each band.

In Figure 11 we compare accuracy against predicted error on polymer-polymer interfaces (first row) and polymer chains (second row). For each, we show a box-and-whiskers plot of accuracy as a function of the error that the model predicts. The reported accuracy metric is interface type-specific (see definitions above). We chose box boundaries at (25%, 75%) percentiles and whisker boundaries at (5%, 95%) percentiles.

In Figure 12 we show accuracy vs predicted error for ligand-protein and ion-protein interfaces, using pocket-aligned RMSD thresholded under various values. The ligands category considers standard ligands only (as described in subsection 3.1 above). The ions category includes both metals and non-metals, with metals representing about 90% of clusters.

4. Conclusion

In this note, we describe our progress on the next generation of AlphaFold model, one that is capable of predicting the joint 3D structure of proteins, nucleic acids, small molecules, ions, and modified residues. Through a series of quantitative benchmarks, we set a new accuracy bar for protein-ligand structure prediction far beyond docking, show state-of-the-art accuracy on protein-DNA and protein-RNA interface structure prediction, and even improve the accuracy of protein-protein interfaces compared to previous AlphaFold models. The inclusion of ions and modified residues allows this model to now represent the majority of PDB structures to a high degree of accuracy, and with a sufficiently calibrated prediction confidence. We presented a range of example predictions, including highly novel and scientifically-relevant complexes, exemplifying the potential for near-term application of this model to challenging domains like therapeutic design. While the work presented is still under active development, we believe it shows that computational structural prediction of general biomolecules is achievable by machine learning models like AlphaFold and opens up many exciting future avenues of research towards understanding biology.

A. Evaluation Process

A.1. Recent PDB Evaluation Set

The recent PDB evaluation set construction started by taking all 10,192 PDB entries released between 2022-05-01 and 2023-01-12, a date range falling after any data in our training set. Each entry in the date range was expanded from the asymmetric unit to Biological Assembly 1, then two filters were applied:

- Filtering to non-NMR entries with resolution better than 4.5 Å, leaving 9,636 complexes.
- Filtering to complexes with less than 5,120 tokens under our tokenization scheme (see section 2), leaving 8,856 complexes.

For each complex we generated a list of all individual entities and a list of all entity pairs where the minimal distance between heavy atoms in the two entities was less than 5 Å and at least one entity in the pair was a polymer. The procedures for determining homology and clusters for these entities and interfaces are described in later sections.

Predictions on the recent PDB set were made on the full post-assembly complex, but crystallisation aids were removed from the complex for prediction and scoring, along with all bonds for structures with homomeric subcomplexes lacking the corresponding homomeric symmetry.

Not every entity and interface was scored:

- Peptide-peptide interfaces, peptide monomers and modified residues within peptides (where a peptide here is defined as a protein with less than 16 residues) were not included in scoring as their homology to the training set was not determined.
- The system can predict other entities like DNA/RNA hybrids, Peptide Nucleic Acids (PNA) and (D) polypeptides, but these entities and interfaces involving them were not scored as they are too rare to get meaningful results on.
- Eight structures were removed from the test set as matching predicted chains to ground truth chains took too long, due to large numbers of individual entities in the structure.
- Four structures were removed from the set for technical reasons - three where all chains had fewer than 4 residues and one with bad metadata in the source file.
- Eleven ligand-protein or ion-protein interfaces failed to score due to RMSD calculation errors.

A.2. Low Homology Subset

The model was trained on structures in the Protein Data Bank (PDB) released before September 30, 2021. In addition to the full evaluation set described in the section above we create a "low homology" subset that is filtered on homology to this training set.

Evaluation is done either on individual chains, or on specific interfaces extracted from the full complex prediction. For intra-chain metrics, we keep polymers that have less than 40% sequence identity to the training set. Here we define sequence identity as the percent of residues in the evaluation set chain that are identical to the training set chain. For interface metrics the following filters are applied:

- Polymer-polymer interfaces: If both polymers have greater than 40% sequence identity to two chains in the same complex in the training set, then this interface is filtered out.

- Polymer-ligand interfaces: If there exists a complex in the training set with a chain that has greater than 40% sequence identity to the polymer and a chain that has greater than 0.5 Tanimoto similarity to the ligand, this interface is filtered out. We only use this low homology measure for bonded ligands and glycans in subsection 3.4 – the main ligands results in subsection 3.1 use a different approach described there.
- Peptide-polymer: For interfaces to a peptide (<16 residues), the similarity of the non-peptide entity has to be novel (less than 40% sequence identity to anything in the training set).

A.3. Clustering

The evaluation data was clustered to allow for redundancy reduction, based on a 40% sequence similarity clustering of all polymer chains. When assigning a cluster ID to an interface:

- Polymer-polymer interfaces are given a cluster ID of (polymer1_cluster, polymer2_cluster).
- Polymer-ligand interfaces are given a cluster ID of the polymer_cluster only.
- Polymer-modified_residue interfaces are given a cluster ID (polymer_cluster, CCD-code).

A.4. Evaluation Metrics

The evaluation procedure compares a predicted structure to the corresponding ground truth structure. If the complex contains multiple identical entities, the optimal assignment (maximising LDDT) of the predicted units to the ground truth units is found by either an exhaustive search over all permutations (for groups up to 8 members) or a simulated annealing optimization (for larger groups). After the chain assignment is found, the 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 (Landrum et al., 2023).

We measure the quality of the predictions with DockQ, LDDT (local distance difference test) (Mariani et al., 2013) or pocket aligned RMSD (root mean square deviation). 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. 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.

If not stated differently, the 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 in the ground truth structure. In ligand-protein interfaces the Cα atoms within the pocket are used to align the predicted structure to the ground truth structure by least squares rigid alignment, and then RMSD is computed on all heavy atoms of the ligand. In ligand-nucleic chain interfaces all heavy atoms are used for the alignment due to the larger spacing of the backbone atoms and the more rigid structure inside.

We used three categories of evaluation metrics:

1. Complex level metrics (e.g. full complex LDDT score). For these we used the prediction with the highest whole complex confidence over the 5 seeds.
2. Individual entity metrics (e.g. intra-chain LDDT). For these we used the prediction with the highest entity confidence for the entity being evaluated.
3. Interface metrics (e.g. pocket aligned RMSD for ligand-protein interfaces). For these we used the prediction with the highest interface confidence for the interface being evaluated.

A.5. Aggregation of Scores

To avoid the overrepresentation of similar polymer chains or interfaces that were deposited under different PDB codes, we cluster those together (see subsection A.3) and aggregate the individual scores such that each cluster gets the same weight. For mean values this is achieved by first computing the mean value per-cluster and then averaging over clusters. For median values and other percentiles we weight each sample by $1/(N_{c_i})$, where N_{c_i} is the size of the cluster the sample belongs to, and compute a weighted percentile.

A.6. Baselines

As part of our ligand accuracy evaluation, we ran docking onto AlphaFold 2.3 structures with AutoDock Vina 1.1 (Eberhardt et al., 2021; Trott and Olson, 2009), using Gypsum-DL (Ropp et al., 2019) and Propka (Olsson et al., 2011) for ligand and protein preparation. The pocket used for docking onto the AlphaFold structures was determined after a rigid alignment with the ground truth structures. We verified the performance of our docking protocol by recovering the published accuracy on ground truth structures from the PoseBusters set (Buttenschoen et al., 2023). All other reported baseline numbers in the ligand section are from published literature.

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 (Baek et al., 2022). We run the open source RF2NA (DiMaio et al., 2023) with the same multiple sequence alignments (MSAs) as were used for AlphaFold-latest predictions. For comparison between AlphaFold-latest and RF2NA, a subset of our recent PDB evaluation set targets are chosen to meet the RF2NA criteria (<1000 total residues and nucleotides). Also as RF2NA was not trained to predict systems with DNA and RNA, this analysis was limited to targets with only one nucleic acid type Figure 9b. Note, however, that AlphaFold-latest is capable of predicting systems with any combination of protein, DNA, RNA, and ligands (see Figure 1).

As an additional baseline for RNA tertiary structure prediction, in Figure 2c above, we evaluate AlphaFold-latest performance on CASP15 RNA targets that are currently publicly available (R1116/8S95, R1117/8FZA, R1126[1], 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 and AIchemy_RNA2 as the top performing entrant. Both the RF2NA (CASP15 entry BAKER) and AIchemy_RNA2 predictions were downloaded from the CASP website.

During preparation of this manuscript, independent work on RoseTTAFold All-Atom (Krishna et al., 2023) was released that performs structure prediction and protein design across a wide range of biomolecular systems. This system is not available for baselining at the time of writing, but the RoseTTAFold All-Atom paper indicates their accuracy is below specialist predictors in almost all categories.

[1] Downloaded from https://predictioncenter.org/casp15/TARGETS_PDB/R1126.pdb

B. Tables

Table 1 | Ligand exclusion list

144, 15P, 1PE, 2F2, 2JC, 3HR, 3SY, 7N5, 7PE, 9JE, AAE, ABA, ACE, ACN, ACT, ACY, AZI, BAM, BCN, BCT, BDN, BEN, BME, BO3, BTB, BTC, BU1, C8E, CAD, CAQ, CBM, CCN, CIT, CL, CLR, CM, CMO, CO3, CPT, CXS, D10, DEP, DIO, DMS, DN, DOD, DOX, EDO, EEE, EGL, EOH, EOX, EPE, ETF, FCY, FJO, FLC, FMT, FW5, GOL, GSH, GTT, GYF, HED, IHP, IHS, IMD, IOD, IPA, IPH, LDA, MB3, MEG, MES, MLA, MLI, MOH, MPD, MRD, MSE, MYR, N, NA, NH2, NH4, NHE, NO3, O4B, OHE, OLA, OLC, OMB, OME, OXA, P6G, PE3, PE4, PEG, PEO, PEP, PG0, PG4, PGE, PGR, PLM, PO4, POL, POP, PVO, SAR, SCN, SEO, SEP, SIN, SO4, SPD, SPM, SR, STE, STO, STU, TAR, TBU, TME, TPO, TRS, UNK, UNL, UNX, UPL, URE

Table 2 | CCD codes defining glycans

045, 05L, 07E, 07Y, 08U, 09X, 0BD, 0H0, 0HX, 0LP, 0MK, 0NZ, 0UB, 0V4, 0WK, 0XY, 0YT, 10M, 12E, 145, 147, 149, 14T, 15L, 16F, 16G, 16O, 17T, 18D, 18O, 1CF, 1FT, 1GL, 1GN, 1LL, 1S3, 1S4, 1SD, 1X4, 20S, 20X, 22O, 22S, 23V, 24S, 25E, 26O, 27C, 289, 291, 293, 2DG, 2DR, 2F8, 2FG, 2FL, 2GL, 2GS, 2H5, 2HA, 2M4, 2M5, 2M8, 2OS, 2WP, 2WS, 32O, 34V, 38J, 3BU, 3DO, 3DY, 3FM, 3GR, 3HD, 3J3, 3J4, 3LJ, 3LR, 3MG, 3MK, 3R3, 3S6, 3SA, 3YW, 40J, 42D, 445, 44S, 46D, 46Z, 475, 48Z, 491, 49A, 49S, 49T, 49V, 4AM, 4CQ, 4GC, 4GL, 4GP, 4JA, 4N2, 4NN, 4QY, 4R1, 4RS, 4SG, 4UZ, 4V5, 50A, 51N, 56N, 57S, 5GF, 5GO, 5II, 5KQ, 5KS, 5KT, 5KV, 5L3, 5LS, 5LT, 5MM, 5N6, 5QP, 5SP, 5TH, 5TJ, 5TK, 5TM, 61J, 62I, 64K, 66O, 6BG, 6C2, 6DM, 6GB, 6GP, 6GR, 6K3, 6KH, 6KL, 6KS, 6KU, 6KW, 6LA, 6LS, 6LW, 6MJ, 6MN, 6PZ, 6S2, 6UD, 6YR, 6ZC, 73E, 79J, 7CV, 7D1, 7GP, 7JZ, 7K2, 7K3, 7NU, 83Y, 89Y, 8B7, 8B9, 8EX, 8GA, 8GG, 8GP, 8I4, 8LR, 8OQ, 8PK, 8S0, 8YV, 95Z, 96O, 98U, 9AM, 9C1, 9CD, 9GP, 9KJ, 9MR, 9OK, 9PG, 9QG, 9S7, 9SG, 9SJ, 9SM, 9SP, 9T1, 9T7, 9VP, 9WJ, 9WN, 9WZ, 9YW, A0K, A1Q, A2G, A5C, A6P, AAL, ABD, ABE, ABF, ABL, AC1, ACR, ACX, ADA, AF1, AFD, AFO, AFP, AGL, AH2, AH8, AHG, AHM, AHR, AIG, ALL, ALX, AMG, AMN, AMU, AMV, ANA, AOG, AQA, ARA, ARB, ARI, ARW, ASC, ASG, ASO, AXP, AXR, AY9, AZC, B0D, B16, B1H, B1N, B2G, B4G, B6D, B7G, B8D, B9D, BBK, BBV, BCD, BDF, BDG, BDP, BDR, BEM, BFN, BG6, BG8, BGC, BGL, BGN, BGP, BGS, BHG, BM3, BM7, BMA, BMX, BND, BNG, BNX, BO1, BOG, BQY, BS7, BTG, BTU, BW3, BWG, BXF, BXP, BXX, BXY, BZD, C3B, C3G, C3X, C4B, C4W, C5X, CBF, CBI, CBK, CDR, CE5, CE6, CE8, CEG, CEZ, CGF, CJB, CKB, CKP, CNP, CR1, CR6, CRA, CT3, CTO, CTR, CTT, D1M, D5E, D6G, DAF, DAG, DAN, DDA, DDL, DEG, DEL, DFR, DFX, DG0, DGO, DGS, DGU, DJB, DJE, DK4, DKX, DKZ, DL6, DLD, DLF, DLG, DNO, DO8, DOM, DPC, DQR, DR2, DR3, DR5, DRI, DSR, DT6, DVC, DYM, E3M, E5G, EAG, EBG, EBQ, EEN, EEQ, EGA, EMP, EMZ, EPG, EQP, EQV, ERE, ERI, ETT, EUS, F1P, F1X, F55, F58, F6P, F8X, FBP, FCA, FCB, FCT, FDP, FDQ, FFC, FFX, FIF, FK9, FKD, FMF, FMO, FNG, FNY, FRU, FSA, FSI, FSM, FSW, FUB, FUC, FUD, FUF, FUL, FUY, FVQ, FX1, FYJ, G0S, G16, G1P, G20, G28, G2F, G3F, G3I, G4D, G4S, G6D, G6P, G6S, G7P, G8Z, GAA, GAC, GAD, GAF, GAL, GAT, GBH, GC1, GC4, GC9, GCB, GCD, GCN, GCO, GCS, GCT, GCU, GCV, GCW, GDA, GDL, GE1, GE3, GFP, GIV, GL0, GL1, GL2, GL4, GL5, GL6, GL7, GL9, GLA, GLC, GLD, GLF, GLG, GLO, GLP, GLS, GLT, GM0, GMB, GMH, GMT, GMZ, GN1, GN4, GNS, GNX, GP0, GP1, GP4, GPH, GPK, GPM, GPO, GPQ, GPU, GPV, GPW, GQ1, GRF, GRX, GS1, GS9, GTK, GTM, GTR, GU0, GU1, GU2, GU3, GU4, GU5, GU6, GU8, GU9, GUF, GUL, GUP, GUZ, GXL, GXV, GYE, GYG, GYP, GYU, GYV, GZL, H1M, H1S, H2P, H3S, H53, H6Q, H6Z, HBZ, HD4, HNV, HNW, HSG, HSH, HSJ, HSQ, HSX, HSY, HTG, HTM, HVC, IAB, IDC, IDF, IDG, IDR, IDS, IDU, IDX, IDY, IEM, IN1, IPT, ISD, ISL, ISX, IXD, J5B, JFZ, JHM, JLT, JRV, JSV, JV4, JVA, JVS, JZR, K5B, K99, KBA, KBG, KD5, KDA, KDB, KDD, KDE, KDF, KDM, KDN, KDO, KDR, KFN, KG1, KGM, KHP, KME, KO1, KO2, KOT, KTU, L0W, L1L, L6S, L6T, LAG, LAH, LAI, LAK, LAO, LAT, LB2, LBS, LBT, LCN, LDY, LEC, LER, LFC, LFR, LGC, LGU, LKA, LKS, LM2, LMO, LNV, LOG, LOX, LRH, LTG, LVO, LVZ, LXB, LXC, LXZ, LZ0, M1F, M1P, M2F, M3M, M3N, M55, M6D, M6P, M7B, M7P, M8C, MA1, MA2, MA3, MA8, MAB, MAF, MAG, MAL, MAN, MAT, MAV, MAW, MBE, MBF, MBG, MCU, MDA, MDP, MFB, MFU, MG5, MGC, MGL, MGS, MJJ, MLB, MLR, MMA, MN0, MNA, MQG, MQT, MRH, MRP, MSX, MTT, MUB, MUR, MVP, MXY, MXZ, MYG, N1L, N3U, N9S, NA1, NAA, NAG, NBG, NBX, NBY, NDG, NFG, NG1, NG6, NGA, NGC, NGE, NGK, NGR, NGS, NGY, NGZ, NHF, NLC, NM6, NM9, NNG, NPF, NSQ, NT1, NTF, NTO, NTP, NXD, NYT, OAK, OI7, OPM, OSU, OTG, OTN, OTU, OX2, P53, P6P, P8E, PA1, PAV, PDX, PH5, PKM, PNA, PNG, PNJ, PNW, PPC, PRP, PSG, PSV, PTQ, PUF, PZU, QDK, QIF, QKH, QPS, QV4, R1P, R1X, R2B, R2G, RAE, RAF, RAM, RAO, RB5, RBL, RCD, RER, RF5, RG1, RGG, RHA, RHC, RI2, RIB, RIP, RM4, RP3, RP5, RP6, RR7, RRJ, RRY, RST, RTG, RTV, RUG, RUU, RV7, RVG, RVM, RWI, RY7, RZM, S7P, S81, SA0, SCG, SCR, SDY, SEJ, SF6, SF9, SFU, SG4, SG5, SG6, SG7, SGA, SGC, SGD, SGN, SHB, SHD, SHG, SIA, SID, SIO, SIZ, SLB, SLM, SLT, SMD, SN5, SNG, SOE, SOG, SOL, SOR, SR1, SSG, SSH, STW, STZ, SUC, SUP, SUS, SWE, SZZ, T68, T6D, T6P, T6T, TA6, TAG, TCB, TDG, TEU, TF0, TFU, TGA, TGK, TGR, TGY, TH1, TM5, TM6, TMR, TMX, TNX, TOA, TOC, TQY, TRE, TRV, TS8, TT7, TTV, TU4, TUG, TUJ, TUP, TUR, TVD, TVG, TVM, TVS, TVV, TVY, TW7, TWA, TWD, TWG, TWJ, TWY, TXB, TYV, U1Y, U2A, U2D, U63, U8V, U97, U9A, U9D, U9G, U9J, U9M, UAP, UBH, UBO, UDC, UEA, V3M, V3P, V71, VG1, VJ1, VJ4, VKN, VTB, W9T, WIA, WOO, WUN, WZ1, WZ2, X0X, X1P, X1X, X2F, X2Y, X34, X6X, X6Y, XDX, XGP, XIL, XKJ, XLF, XLS, XMM, XS2, XXM, XXR, XXX, XYF, XYL, XYP, XYS, XYT, XYZ, YDR, YIO, YJM, YKR, YO5, YX0, YX1, YYB, YYH, YYJ, YYK, YYM, YYQ, YZ0, Z0F, Z15, Z16, Z2D, Z2T, Z3K, Z3L, Z3Q, Z3U, Z4K, Z4R, Z4S, Z4U, Z4V, Z4W, Z4Y, Z57, Z5J, Z5L, Z61, Z6H, Z6J, Z6W, Z8H, Z8T, Z9D, Z9E, Z9H, Z9K, Z9L, Z9M, Z9N, Z9W, ZB0, ZB1, ZB2, ZB3, ZCD, ZCZ, ZD0, ZDC, ZDO, ZEE, ZEL, ZGE, ZMR

Table 3 | Ions

118, 119, 1AL, 1CU, 2FK, 2HP, 2OF, 3CO, 3MT, 3NI, 3OF, 4MO, 4PU, 4TI, 543, 6MO, AG, AL, ALF, AM, ATH, AU, AU3, AUC, BA, BEF, BF4, BO4, BR, BS3, BSY, CA, CAC, CD, CD1, CD3, CD5, CE, CF, CHT, CO, CO5, CON, CR, CS, CSB, CU, CU1, CU2, CU3, CUA, CUZ, CYN, DME, DMI, DSC, DTI, DY, E4N, EDR, EMC, ER3, EU, EU3, F, FE, FE2, FPO, GA, GD3, GEP, HAI, HG, HGC, HO3, IN, IR, IR3, IRI, IUM, K, KO4, LA, LCO, LCP, LI, LU, MAC, MG, MH2, MH3, MMC, MN, MN3, MN5, MN6, MO, MO1, MO2, MO3, MO4, MO5, MO6, MOO, MOS, MOW, MW1, MW2, MW3, NA2, NA5, NA6, NAO, NAW, NET, NI, NI1, NI2, NI3, NO2, NRU, O4M, OAA, OC1, OC2, OC3, OC4, OC5, OC6, OC7, OC8, OCL, OCM, OCN, OCO, OF1, OF2, OF3, OH, OS, OS4, OXL, PB, PBM, PD, PER, PI, PO3, PR, PT, PT4, PTN, RB, RH3, RHD, RU, SB, SE4, SEK, SM, SMO, SO3, T1A, TB, TBA, TCN, TEA, TH, THE, TL, TMA, TRA, V, VN3, VO4, W, WO5, Y1, YB, YB2, YH, YT3, ZCM, ZN, ZN2, ZN3, ZNO, ZO3, ZR

Table 4 | Standard residues

ALA, ARG, ASN, ASP, CYS, GLN, GLU, GLY, HIS, ILE, LEU, LYS, MET, PHE, PRO, SER, THR, TRP, TYR, VAL, UNK, A, G, C, U, DA, DG, DC, DT, N

References

AlphaFold-Team. AlphaFold 2.3, 12 2022. URL https://github.com/google-deepmind/alphafold/blob/main/docs/technical_note_v2.3.0.md.

M. Baek, R. McHugh, I. Anishchenko, D. Baker, and F. DiMaio. Accurate prediction of nucleic acid and protein-nucleic acid complexes using rosettafoldna. *bioRxiv preprint bioRxiv:10.1101/2022.09.09.507333*, 2022.

M. Buttenschoen, G. M. Morris, and C. M. Deane. Posebusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *arXiv preprint arXiv:2308.05777*, 2023.

K. Chen, Y. Zhou, S. Wang, and P. Xiong. RNA tertiary structure modeling with briq potential in casp15. *Proteins: Structure, Function, and Bioinformatics*, 2022. doi: 10.1002/prot.26574.

F. DiMaio, B. Riley, and A. Morehead. RoseTTAFold2NA, April 2023. URL https://github.com/uw-ipd/RoseTTAFold2NA/releases/tag/v0.2.

J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli. Autodock vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8): 3891–3898, Aug 2021. ISSN 1549-9596. doi: 10.1021/acs.jcim.1c00203.

R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv preprint bioRxiv:10.1101/2021.10.04.463034v2*, 2021.

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

M. Karelina, J. J. Noh, and R. O. Dror. How accurately can one predict drug binding modes using AlphaFold models? *eLife*, aug 2023. doi: 10.7554/elife.89386.1.

J. Kreitz, M. J. Friedrich, A. Guru, B. Lash, M. Saito, R. K. Macrae, and F. Zhang. Programmable protein delivery with a bacterial contractile injection system. *Nature*, 616(7956):357–364, Apr 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-05870-7.

R. Krishna, J. Wang, W. Ahern, P. Sturmfels, P. Venkatesh, I. Kalvet, G. R. Lee, F. S. Morey-Burrows, I. Anishchenko, I. R. Humphreys, R. McHugh, D. Vafeados, X. Li, G. A. Sutherland, A. Hitchcock, C. N. Hunter, M. Baek, F. DiMaio, and D. Baker. Generalized biomolecular modeling and design with rosettafold all-atom. *bioRxiv preprint bioRxiv:10.1101/2023.10.09.561603*, 2023.

A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, and J. Moult. Critical assessment of methods of protein structure prediction (CASP) – round XV. oct 2023. doi: 10.22541/au.169658651.11658182/v1. URL https://doi.org/10.22541%2Fau.169658651.11658182%2Fv1.

G. Landrum, P. Tosco, B. Kelley, Ric, D. Cosgrove, sriniker, gedeck, R. Vianello, NadineSchneider, E. Kawashima, D. N, G. Jones, A. Dalke, B. Cole, M. Swain, S. Turk, AlexanderSavelyev, A. Vaucher, M. Wójcikowski, I. Take, D. Probst, K. Ujihara, V. F. Scalfani, guillaume godin, J. Lehtivarjo, R. Walker, A. Pahl, F. Berenger, jasondbiggs, and strets123. rdkit/rdkit: 2023_03_3 (q1 2023) release, Aug. 2023.

Y. Lim, L. Tamayo-Orrego, E. Schmid, Z. Tarnauskaite, O. V. Kochenova, R. Gruar, S. Muramatsu, L. Lynch, A. V. Schlie, P. L. Carroll, G. Chistol, M. A. M. Reijns, M. T. Kanemaki, A. P. Jackson, and J. C. Walter. In silico protein interaction screening uncovers donson's role in replication initiation. *Science*, 381(6664):eadi3448, 2023. doi: 10.1126/science.adi3448.

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

S. Mosalaganti, A. Obarska-Kosinska, M. Siggel, R. Taniguchi, B. Turoňová, C. E. Zimmerli, K. Buczak, F. H. Schmidt, E. Margiotta, M.-T. Mackmull, W. J. H. Hagen, G. Hummer, J. Kosinski, and M. Beck. Ai-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science*, 376(6598):eabm9506, 2022. doi: 10.1126/science.abm9506.

M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen. Propka3: Consistent treatment of internal and surface residues in empirical pka predictions. *Journal of Chemical Theory and Computation*, 7(2):525–537, Feb 2011. ISSN 1549-9618. doi: 10.1021/ct100578z.

P. J. Ropp, J. O. Spiegel, J. L. Walker, H. Green, G. A. Morales, K. A. Milliken, J. J. Ringe, and J. D. Durrant. Gypsum-dl: an open-source program for preparing small-molecule libraries for structure-based virtual screening. *Journal of Cheminformatics*, 11(1):34, May 2019. ISSN 1758-2946. doi: 10.1186/s13321-019-0358-3.

V. Scardino, J. I. Di Filippo, and C. N. Cavasotto. How good are AlphaFold models for docking-based virtual screening? *iScience*, 26(1):105920, 2023. ISSN 2589-0042. doi: 10.1016/j.isci.2022.105920.

O. Trott and A. J. Olson. AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, June 2009. doi: 10.1002/jcc.21334.

P. Xiong, R. Wu, J. Zhan, and Y. Zhou. Pairing a high-resolution statistical potential with a nucleobase-centric sampling algorithm for improving RNA model refinement. *Nature Communications*, 12(1):2777, May 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-23100-4.