nature chemical biology

Perspective

https://doi.org/10.1038/s41589-024-01638-w

# The power and pitfalls of AlphaFold2 for structure prediction beyond rigid globular proteins

Received: 22 May 2023

Accepted: 29 April 2024

Published online: 21 June 2024

Check for updates

Vinayak Agarwal [1,2] ✉ & Andrew C. McShan [1] ✉

Artificial intelligence-driven advances in protein structure prediction in recent years have raised the question: has the protein structure-prediction problem been solved? Here, with a focus on nonglobular proteins, we highlight the many strengths and potential weaknesses of DeepMind's AlphaFold2 in the context of its biological and therapeutic applications. We summarize the subtleties associated with evaluation of AlphaFold2 model quality and reliability using the predicted local distance difference test (pLDDT) and predicted aligned error (PAE) values. We highlight various classes of proteins that AlphaFold2 can be applied to and the caveats involved. Concrete examples of how AlphaFold2 models can be integrated with experimental data in the form of small-angle X-ray scattering (SAXS), solution NMR, cryo-electron microscopy (cryo-EM) and X-ray diffraction are discussed. Finally, we highlight the need to move beyond structure prediction of rigid, static structural snapshots toward conformational ensembles and alternate biologically relevant states. The overarching theme is that careful consideration is due when using AlphaFold2-generated models to generate testable hypotheses and structural models, rather than treating predicted models as de facto ground truth structures.

DeepMind's AlphaFold2 (AF2) has revolutionized structural biology with its deep learning algorithm that enables accurate prediction of three-dimensional (3D) protein structures from only the target amino acid sequence, potentially solving the half-a-century-old protein structure-prediction problem: how to predict 3D structures from only sequence information[1–4]. AF2 has opened the door to understanding protein folds, structures, interactions and function at organismal levels through modeling of 98.5% of the human proteome[5]. One common critique of AF2 is that it requires substantial computational resources to run the software locally (up to 3 TB of disk space and a modern NVIDIA graphics processing unit with gigabytes of memory). Several efforts have alleviated these limitations, including the AlphaFold Protein Structure Database[6,7], which houses over 200 million pre-run AF2 predictions, as well as ColabFold[8] and OpenFold[9], which allow users to run a modified AF2 protocol on open-access servers in minutes. These platforms have allowed the public, industry and academics without computational resources to model and analyze structures of their favorite target using AF2 with just a few clicks of a button. Another critique of AF2 concerns whether it has truly solved the protein structure-prediction problem. Several groups have proposed that AF2 has only learned how to estimate 3D structures using patterns extracted from known folds in the Protein Data Bank (PDB) and coevolutionary information between residues rather than the underlying physical and chemical basis of protein folding[3,10–12]. This is strictly true because current versions of AF2 do not use energy functions that seek to identify native-like protein conformations, unlike its competitor Rosetta[13]. Others suggest that AF2's algorithm may have indirectly learned a similar function[14]. Finally, some critics question

[1]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA, USA. [2]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA. ✉e-mail: vagarwal@gatech.edu; andrew.mcshan@chemistry.gatech.edu

https://doi.org/10.1038/s41589-024-01638-w

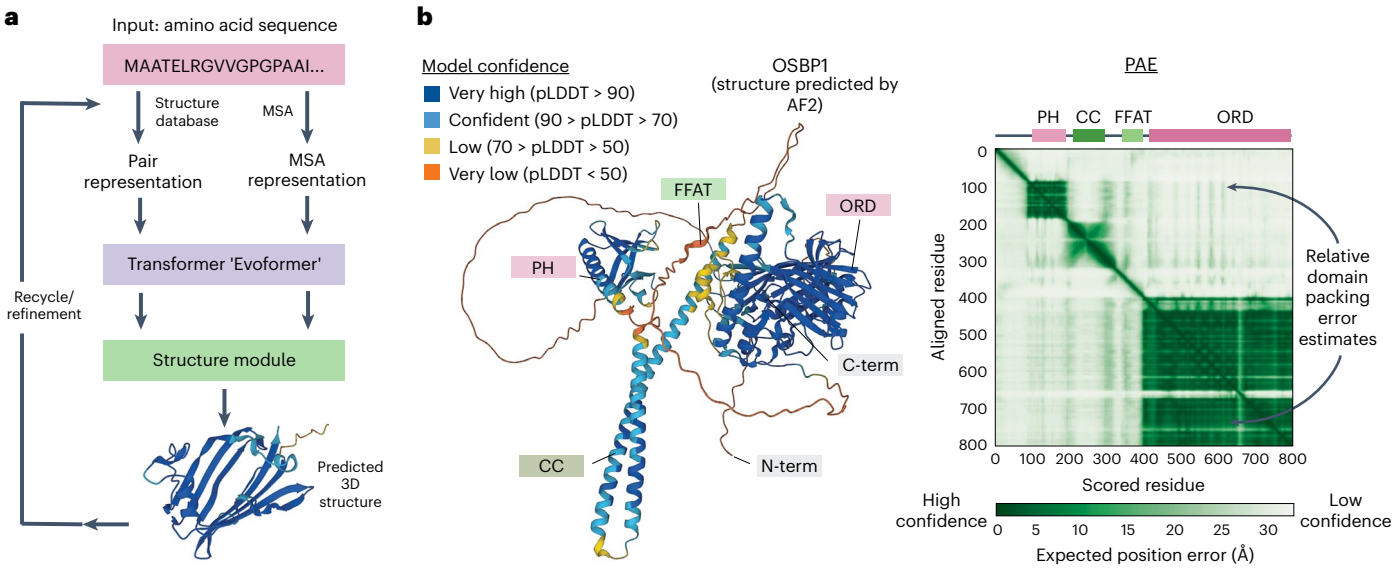


**Fig. 1 | Overview of AF2. a**, The general workflow for an AF2 prediction is shown (derived from Jumper et al.[1]). The input is the primary amino acid sequence. The AF2 model of ganglioside GM2 activator protein (https://alphafold.ebi.ac.uk/entry/P17900) is shown. The model is colored based on pLDDT values. **b**, Left, AF2 prediction for OSBP1 (https://alphafold.ebi.ac.uk/entry/P22059) is shown. The AF2 model is colored based on pLDDT values. The domains of the protein are noted: PH, FFAT, ORD and CC. Term, terminus. Right, AF2 PAE metrics show the predicted relative position error for each residue in the sequence, with low-confidence values in white and high-confidence values in green. The domains of OSBP1 have been manually annotated on the PAE graph.

the accuracy of the standard implementation of AF2 against different types of nonglobular molecular targets, which could limit its potential applications[15,16]. Overwhelming evidence suggests that machine learning software like AF2, RoseTTAFold, ESMFold, and related approaches are the best and most accurate answer to the structure-prediction problem to date[1,17–19].

## AF2's artificial intelligence-driven revolution

The basic workflow of AF2 is outlined in Fig. 1a. Users input the primary amino acid sequence of the target protein as one-letter code in FASTA format. When more than one input sequence is provided, AlphaFold-Multimer or AF2Complex is used[20,21]. Lower and upper limits for input sequence lengths are defined by difficulties in generating reliable multiple-sequence alignments (MSAs) for short (less than ten amino acids) sequences and graphic processing and/or memory issues for long (>3,000 amino acids) sequences, respectively. Protein sequences can be obtained from annotated public databases, such as UniProt. The full details of the AF2 workflow have been discussed previously[1] but are briefly outlined below. Using the input sequence(s), AF2 first queries several databases to construct a pair representation and an MSA representation of the target. The pair representation is a matrix of pairwise interactions between amino acids that are likely to be spatially related (that is, close to each other in space). The MSA representation is a collection of sequences that are evolutionarily related to the target sequence and provides mutational covariance information used by AF2. The pair and MSA representations are then passed through the Evoformer, a neural network block that exchanges information within the MSA and pair representations to establish spatial and evolutionary relationships. Next, the structural module parses information from the Evoformer to convert the representations into a 3D protein structure. The entire process undergoes several rounds of iterative recycling to produce the final refined models. For each output, AF2 generates a per-residue confidence score stored in the *B*-factor column of the model coordinate file (.pdb, .mmCIF or related formats), the pLDDT score, which ranges from 0 to 100, with higher values assigned higher confidence in the model[1,22] (Fig. 1b). AF2 also generates a PAE matrix, which evaluates the relative orientation and position of different parts (that is, domains) of the model[6]. Higher PAE values correspond to lower confidence for the relative position and orientation of two parts of the protein in the model. Users should be especially careful to not assign biological or structural relevance to regions with low pLDDT (<70) or high PAE (>5 Å) values[5,23]. However, as discussed below, high pLDDT or low PAE metrics, indicating high confidence in the prediction, do not promise agreement with native protein conformations but instead estimate a likelihood for local and global coordinate positions and/or orientations.

An example AF2 model of oxysterol-binding protein 1 (OSBP1), a lipid transfer protein, obtained from the AlphaFold Protein Structure Database is shown in Fig. 1b. The pLDDT values plotted onto the model highlight that the pleckstrin homology (PH), coiled-coil (CC) and OSBP-related ligand-binding (ORD) domain structure are assigned very high to high predicted confidence, while the phenylalanines in an acidic tract motif (FFAT) domain are predicted with very low confidence. The PAE graph reveals that the model has low confidence with respect to the relative placement of PH, CC, FFAT and ORD domains with respect to each other.

## A critical evaluation of AF2's applications

There are several open areas of research concerning AF2 in the context of its biological and therapeutic applications: first, accuracy evaluations of AF2 models relative to different types of protein folds present in the PDB, especially for new structures as they are released; second, expanding the types of systems that AF2 can be applied to, either through benchmarking the default AF2 pipeline on new types of targets or through modifications in the AF2 protocol. A recent structural biology community assessment reports that, on average, AF2 generates models with quality near that of experimental structures across diverse target folds and applications[17]. These types of studies with good reason affirm AF2's utility but may give the impression that AF2 is without limitations. Below, we summarize several potential applications of AF2 and provide examples in which the predicted model deviates from the experimental structure. Cases in which AF2's performance is compromised are especially important to help us understand its limitations and provide opportunities to refine its deep learning-based algorithm in future iterations. These deviations can be broadly categorized into cases with (1) inaccurate secondary, tertiary and/or quaternary

https://doi.org/10.1038/s41589-024-01638-w

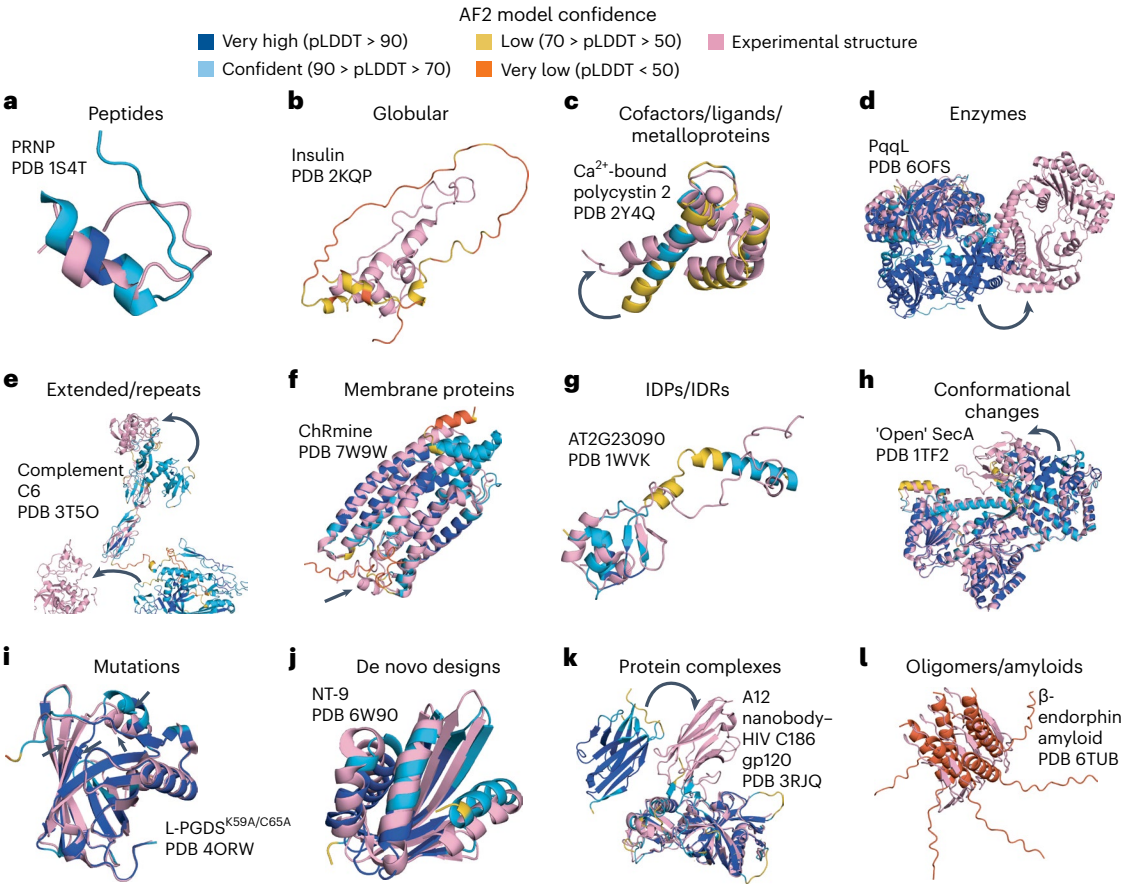


**Fig. 2 | Example applications of AF2 predictions that deviate from the experimental structure.** Superpositions of the AF2 model and experimental structure for several classes of peptides and proteins are shown. AF2 models are colored according to pLDDT values, with overlaid experimental structures colored in pink. The Protein Data Bank (PDB) IDs of the experimental structures used for comparison are noted. **a**–**l**, AF2 models were fetched from the AlphaFold Protein Structure Database or derived from the literature: PRNP (https://alphafold.ebi.ac.uk/entry/P23907) (**a**), insulin (https://alphafold.ebi.ac.uk/entry/P01308) (**b**), polycystin 2 (https://alphafold.ebi.ac.uk/entry/Q13563) (**c**), PqqL (https://alphafold.ebi.ac.uk/entry/P31828) (**d**), complement component C6 (https://alphafold.ebi.ac.uk/entry/P13671) (**e**), ChRmine (**f**), AT2G23090 (https://alphafold.ebi.ac.uk/entry/O64818) (**g**), SecA (https://alphafold.ebi.ac.uk/entry/P28366) (**h**), lipocalin-type PGDS (L-PGDS)$^{K59A/C65A}$ (**i**), NT-9 (**j**), A12 nanobody–HIV C186 gp120 complex (**k**), and β-endorphin amyloid fibril (**l**). Black arrows denote areas where the AF2 model deviates from the experimental structure. It is important to note that predicted structures, gleaned from literature or the AlphaFold structure database, have been generated using different versions of AF2 software and/or with different input parameters and thus cannot be directly compared.

structure in regions where AF2 predicts low to moderate confidence, (2) inaccurate structure in regions where AF2 predicts high model confidence, (3) correct backbone structure but incorrect fine details (that is, side chain rotamer placements), and (4) correct backbone structure for individual domains but inaccurate placement of domains relative to each other (Fig. 2a–l). For cases (1) and (4), low confidence in pLDDT scores and PAE graphs alerts users to interpret structures with caution, which is not immediately clear in cases (2) and (3).

The success of AF2 in predicting protein structures begs the question as to whether it can also accurately predict peptide structures (or, in some cases, lack of a well-defined structure). Peptide structure prediction poses additional challenges given that the benchmark set used to train AF2 excluded peptides, the difficulty in generating robust MSAs for short sequences and observations that many peptides exist in solution as conformational ensembles rather than a single static conformation[1,24,25]. McDonald et al.[25] performed a benchmark of 588 peptides, revealing that AF2 predicts many α-helical and β-hairpin peptide structures with surprising accuracy. However, AF2 was challenged by mixed secondary structure membrane and soluble peptides, such as the prion protein PRNP[25] (Fig. 2a). It was also shown that the best-ranked AF2 models (selected on the basis of high pLDDT score) often did not exhibit the lowest $C_\alpha$ root mean square deviation relative to the experimental structure, suggesting that the pLDDT metric used by AF2 to assess protein models is not optimal for classification of peptide conformations[25]. In a separate study, Tsaban et al.[24] showed that AF2 can be adapted to accurately model peptide–protein complexes irrespective of peptide length, although the results seemed biased toward helical structures and peptides that do not undergo large structural rearrangements upon binding. New methods are fine-tuning AF2-based pipelines for specific types of peptide–protein complexes (that is, peptide–major histocompatibility complex)[26]. These studies provide compelling evidence that AF2 can be applied across peptides, proteins and peptide–protein complexes, albeit with several limitations and caveats. Refinement of AF2-derived models with NMR-derived restraints, such as chemical shift values, torsion angles, residual dipolar couplings (RDCs) and nuclear Overhauser effect (NOE) data, could help improve accuracy of peptide modeling[15].

To date, the most common types of protein folds benchmarked in AF2 assessments are globular and extended or repeat proteins[1,5,17]. NMR structural ensembles offer a unique validation metric to assess the accuracy of predicted AF2 models, as AF2 was trained on a subset of the PDB that excluded NMR data[1,27–30]. While AF2 performs exceptionally well on these types of folds on average, Fowler et al.[28] revealed that NMR ensembles can be more accurate than static AF2 models for dynamic proteins. As an example, the AF2 model of insulin deviates substantially from its experimental NMR structure (Fig. 2b), potentially due to the

inability of AF2 to orient cysteine pairs for disulfide bond formation[31]. Another consideration is that AF2 models of many globular proteins, especially enzymes and metalloproteins, lack functionally relevant cofactors, prosthetic groups or ligands. The authors of AF2 note that, because it is trained on both apo and holo structures from the PDB, models may still be consistent with the expected structure in the presence of ligands or cofactors despite their absence in the AF2 workflow[1,5]. However, whether the modeled structure resembles the apo or holo form of the protein is not immediately clear from analysis of pLDDT scores or PAE graphs[32]. Furthermore, deviations of AF2 models from experimental structures also occur when cofactors, prosthetic groups or ligands induce structural changes, either locally or allosterically. For example, the NMR structure of $Ca^{2+}$-bound polycystin 2 deviates from the AF2 model, potentially due to conformational changes upon $Ca^{2+}$ binding (Fig. 2c). Likewise, the AF2 model of the zinc protease PqqL deviates from the open, highly extended conformation determined by X-ray crystallography (Fig. 2d). New algorithms, such as AlphaFill, are actively being developed that could improve AF2 structure prediction and refinement for cofactor-, prosthetic group- or ligand-bound proteins[33]. These modifications will enable AF2 to identify new therapeutic candidates[34]. AF2 may also exhibit difficulties in structure prediction for extended proteins or proteins with repeat elements[35]. In the case of the extended complement C6 protein, AF2 predicts the structure of individual domains well but deviates in the placement of domains relative to each other (Fig. 2e). For large macromolecules, users may be able to estimate the likelihood that AF2 correctly placed domains relative to each other by visualization of confidence scores in the PAE graph. However, it is important to remember that the PAE values are only confidence estimates. Furthermore, the accuracy of PAE graphs for interdomain prediction has not been as extensively benchmarked as for intradomain contacts[36].

Evaluation of membrane protein structure is another important application of AF2 (ref. 5,37). Benchmarking AF2 against membrane proteins represents a challenge, as the membrane environment, which includes lipids and other proteins, is not directly considered by current versions of AlphaFold[38]. Furthermore, membrane proteins represent less than 3% of total structures in the PDB[15], meaning that the training set used by AF2 was highly biased toward soluble proteins[39]. Hegedűs et al.[38] benchmarked several membrane proteins not included in the original AF2 training set and concluded that, on average, AF2 predicts transmembrane proteins as well as soluble proteins. However, the authors note two important limitations. First, AF2 models with transmembrane region lengths corresponding to nonphysiological membrane thickness values can exhibit very high pLDDT scores (high model confidence), suggesting that pLDDT scores alone are not sufficient to select native membrane protein conformations. Second, AF2 performs poorly for targets embedded in membrane thickness outside the range of 15–35 Å as well as targets with novel features not commonly present in the PDB. In agreement with these findings, Azzaz et al.[40] have shown the difficulty of AF2 in modeling membrane proteins owing to 'epigenetic' factors (that is, lipid environment, co-receptor-induced structural changes, post-translational modifications) that control protein structure beyond the amino acid sequence. As an example, while the AF2 model of the channelrhodopsin ChRmine captures its overall fold, the modeled N-terminal region and extracellular loops deviate from its experimental high-resolution cryo-EM structure (Fig. 2f), likely due to ChRmine's unique covalent Schiff base feature[41]. It will be imperative to evaluate AF2 against membrane protein structures as they become more readily available as the result of high-resolution cryo-EM and advances in NMR spectroscopy.

Another unknown is how AF2 performs on intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs)[27,42]. IDPs represent a challenge for AF2 because it is difficult to identify evolutionary constraints from MSAs of IDPs and IDRs due to sequence hypervariability. In addition, like peptides, IDPs and IDRs are best thought of as sampling diverse conformational ensembles rather than a single static conformation[42]. Preliminary studies suggest that the majority of targets with very low confidence score (pLDDT < 50) assigned by AF2 are likely to be IDPs or IDRs rather than well-folded structures that AF2 fails to predict[42–44]. However, for many targets, AF2 models with predicted disorder may not be relevant for structure and function analysis other than for assigning the likelihood for conformational heterogeneity[27]. As an example, the NMR structure of IDR-containing protein AT2G23090 deviates from the AF2 model despite the confident pLDDT score (Fig. 2g). A study by Ruff et al.[42] showed that the radius of gyration values of IDPs or IDR-containing proteins calculated using static AF2 models substantially deviates from those experimentally obtained by SAXS. Future benchmarks should continue to evaluate AF2 against panels of IDPs and IDR-containing proteins using novel critical assessment of protein intrinsic disorder targets[44]. Efforts are also underway to establish whether AF2 can be used to predict alternative conformations or conformational ensembles of folded proteins (discussed in detail below). Several groups have suggested that the default AF2 pipeline has difficulty in modeling alternative conformations[45]. For example, AF2 fails to predict the 'open' activated conformation of the ATPase SecA (Fig. 2h). Interestingly, several groups have shown that modifications of AF2 have the potential to generate models that substantially deviate from each other, allowing for sampling of conformational landscapes. A study by del Alamo et al.[46] modified the AF2 pipeline by reducing the number of recycles and restricting the depth of randomly subsampled MSAs to sample functionally relevant alternative conformations of transporters and G-protein-coupled receptors. Similarly, Wayment-Steele et al.[47] found that clustering MSAs by sequence similarity enables AF2 to sample known alternative states of KaiB, RfaH and mitotic arrest deficient 2 (MAD2). Further benchmarking of modified AF2 protocols against IDPs, IDR-containing proteins and alternative conformation is required to establish protein prediction strengths and limitations for those of systems[43,48].

There are several other challenging structural modeling problems in biology and therapeutics that AF2 is tasked with. One of the most sought-after applications of AF2 is predicting the effect of mutations on protein structure and/or stability[17]. The AF2 authors note that "AlphaFold has not been trained or validated for predicting the effect of mutations". In support of this, studies have reported an inability of AF2 to predict the effects of mutations on protein structure and stability[14,49,50], which may be due to a training bias on stable structures or an inability to extract signal from small mutations through MSAs. As an example, structural perturbations induced by the K59A/C65A double mutation in prostaglandin D synthase (PGDS) are not accurately captured by AF2 (Fig. 2i). A recent adaption of AlphaFold, AlphaMissense, does not explicitly determine the structural effects of a mutation on a protein but provides the probability of a missense variant being pathogenic[51]. Other groups have suggested that developing AF2 workflows that are less dependent on MSAs could be beneficial[14].

Another challenge is modeling of novel 3D folds that are either completely absent or not commonly represented in the PDB, such as de novo designed proteins. In these cases, AF2 has not been fully trained on novel topologies, which are not commonly found in the PDB. Furthermore, extraction of coevolutionary information from MSAs using de novo designed targets may be difficult, as the amino acid sequences of de novo designed proteins deviate from naturally observed sequences. For some, this is the ultimate test of whether AlphaFold may have solved the protein structure-prediction problem. Interestingly, Moffat et al.[52] showed that AF2 performs well on the de novo designed proteins Top7, Peak6, Foldit1 and Ferredog-Diesel. Slight deviations in tertiary structure are noted, such as for the nuclear transport factor 2-derived de novo designed protein NT-9, but the overall structure is well described (Fig. 2j). For targets without known homologs, such as computationally designed proteins, increasing the

number of recycling iterations can improve the quality of the prediction[8]. An inverted version of AlphaFold, called AlphaDesign, has been used for de novo protein design with some success[53].

Evaluation of protein–protein interactions, including oligomerization, is another major potential application of AF2 that continues to be explored[17]. Yin et al.[55,54] benchmarked 152 heterodimeric protein complexes, revealing that AF2 and AlphaFold2-Multimer had a 51% success rate. The authors note that AF2 had difficulty modeling antibody–protein complexes, such as the A12 nanobody–human immunodeficiency virus (HIV) gp120 complex (Fig. 2k). A separate study by Bryant et al.[56] reported a 63% success rate for heterodimeric complexes. Both studies suggest that a robust MSA coevolutionary signal is required for accurate complex modeling. Preliminary reports also suggest that AF2 may also be able to predict oligomeric states of proteins and amyloids[17,57]. However, care must be taken when interpreting predictions, as highlighted by the incorrect AF2 model of the β-endorphin amyloid fiber relative to its experimental solid-state NMR structure[58] (Fig. 2l).

## Evaluation metrics and model reliability

As stated above, AlphaFold provides error categorizations in the form of pLDDT scores and PAE values to estimate the confidence of its predictions and to evaluate overall model quality and reliability. For the majority of globular proteins, AF2 provides accurate, reliable models with high pLDDT (>70) or low PAE (<5 Å) values highlighting confidence in the prediction of the position of the atomic coordinates, which match experimentally determined native structures[5,17,23,59]. In other cases, if the 'best' AF2 model exhibits many residues with low pLDDT (<70) or high PAE (>5 Å) values, the likelihood that the backbone structure matches the native conformation is very low and the model cannot be reliably interpreted. Previous analysis suggests that AF2 predicts on average ~50% of residues across all proteins with high confidence[17,59]. Users can attempt to increase model quality (better pLDDT and PAE values) by generating a series of predictions with different parameters (number of recycles, number of random seeds, number of ensembles)[1] or by integration with experimental data[60]. However, cases in which the AF2 evaluation metrics are good but the model does not match experimental structure (Fig. 2a,d,e,k) suggest that care must be taken in blind faith in pLDDT and PAE metrics. The most dramatic case is when AF2 provides excellent evaluation metrics despite complete disagreement of the model's backbone structure with an experimental structure. Terwilliger et al.[61] estimate that ~10% of residues predicted by AF2 with high confidence deviate from the backbone by more than 2 Å from native conformations observed in experimental structures. There are also cases in which AF2 generates models with high confidence where the backbone structure is correct but fine details, such as side chain rotamer placement, are lacking. Jumper et al.[1] note that a rotamer is generally classified as correct if the predicted torsion angle is within 40° of the experimental torsion angle, which is correlated with pLDDT scores >90. However, as noted by several groups, high pLDDT at a specific residue does not always indicate that the correct rotameter has been modeled[23]. Cases can also exist in which AF2 predicts the correct backbone structure for individual domains but misplaces domains relative to each other, which should be recognizable in the output PAE matrices. Several groups have reported cases of AF2 models with low PAE values (<5 Å) that deviated from experimental data[62,63]. While no precise mechanism exists to identify these cases, some groups have used molecular dynamics simulations to further evaluate the stability and quality of AF2 models[64,65]. Some groups have used MD simulations to suggest that pLDDT and PAE metrics provide information on dynamics and disorder[42,65]. However, other reports have compared pLDDT scores with crystallographic *B* factors to suggest that AF2 confidence metrics are unable to provide direct information on local flexibility[66]. The determinants driving cases in which AF2 models are associated with high confidence but deviate from experimental structure are currently unknown and should be thoroughly evaluated in future studies, especially in the context of nonglobular proteins, toward quantitatively defining limits of AlphaFold's evaluation and error-categorization metrics. Updated and refined approaches for error categorization may provide better methods for model quality assessment relative to pLDDT and PAE metrics[14,67].

## Integration of AF2 models with experimental data

In cases where no experimental data are available (that is, in vitro recombinant protein production or in situ characterization is not possible), insights into the structure and function of proteins may be primarily guided by AF2 predictions supplemented with molecular dynamics simulations to further evaluate model stability[64,65]. In cases where recombinant protein can be prepared in the milligram quantities required for biophysical characterization, AF2 models can be integrated with experimental data, typically in the form of SAXS, NMR, X-ray crystallography and cryo-EM (Fig. 3a–d). Here, experimental results are directly compared and contrasted against a series of AF2 models to evaluate which prediction, if any, adequately fits the data. AF2 models are increasingly used as initial templates to fit experimental data. The models subsequently undergo further refinement in an iterative fashion to match data toward generation of data-driven structural models. Another possibility is the use of implicit experimental data to guide and restrain AF2 predictions (that is, AF2 models are refined to best fit experimental data)[60]. The integration of AF2 models with experiments is especially useful for cases in which template structures or homologous models are lacking. The use of AlphaFold models in structure-determination protocols has been shown to reduce the time and effort required relative to ab initio model building[68–70].

As one example, theoretical SAXS profiles for a series of AF2 models can be predicted from the 3D coordinates and directly compared with experimental SAXS data in the form of $P(r)$ versus $r$ or log ($I_q$) versus $q$ plots, where $\chi^2$ values provide a goodness-of-fit measure for AF2 models relative to the solution-state structure, which is time and ensemble averaged in SAXS[71–73]. The best-matching AF2 model is fitted into the experimental SAXS-derived envelope using a variety of software for further refinement[73] (Fig. 3a). Preliminary comparison of SAXS-derived versus AlphaFold calculated $P(r)$ curves revealed that, for many cases, a static AF2 model does not adequately describe solution-state structures[42,72]. Recent methods have shown that fitting of SAXS data substantially improves when an ensemble of AlphaFold-predicted structures is used rather than a static AlphaFold model[71], highlighting the importance of integrating AlphaFold models with experimental data. An important caveat is that one must be wary of overfitting AF2 models to SAXS envelopes, especially for lower-resolution data[74]. Typically, $\chi^2$ values of less than one are indicative of overfitting, and additional strategies such as the combination of $V_c$, $Qr$, $X2_{free}$ and $R_{sas}$ metrics have been proposed as more robust evaluation metrics[75].

AF2 models have also been increasingly used during molecular replacement and phasing of X-ray diffraction data obtained from protein crystals[76–79] (Fig. 3b). Standard molecular replacement strategies require 3D coordinates of a template or homologous structure and work best when the template is <2 Å $C_\alpha$ root mean square deviation from the target structure[80]. Recently, an AF2-integrated iterative procedure for molecular replacement has been developed in which AlphaFold models are used during the initial structure-solution cycle, followed by data-guided cycles of AlphaFold structure prediction and model rebuilding[60,69]. This iterative procedure works extremely well as demonstrated in a benchmark in which 187 of 215 structures were solved by AlphaFold-guided molecular replacement; the success was shown to be dependent on high confidence scores associated with the AlphaFold prediction[69]. The use of AlphaFold models in molecular replacement can be further enhanced by downweighting or removing low-confidence regions[23].

https://doi.org/10.1038/s41589-024-01638-w

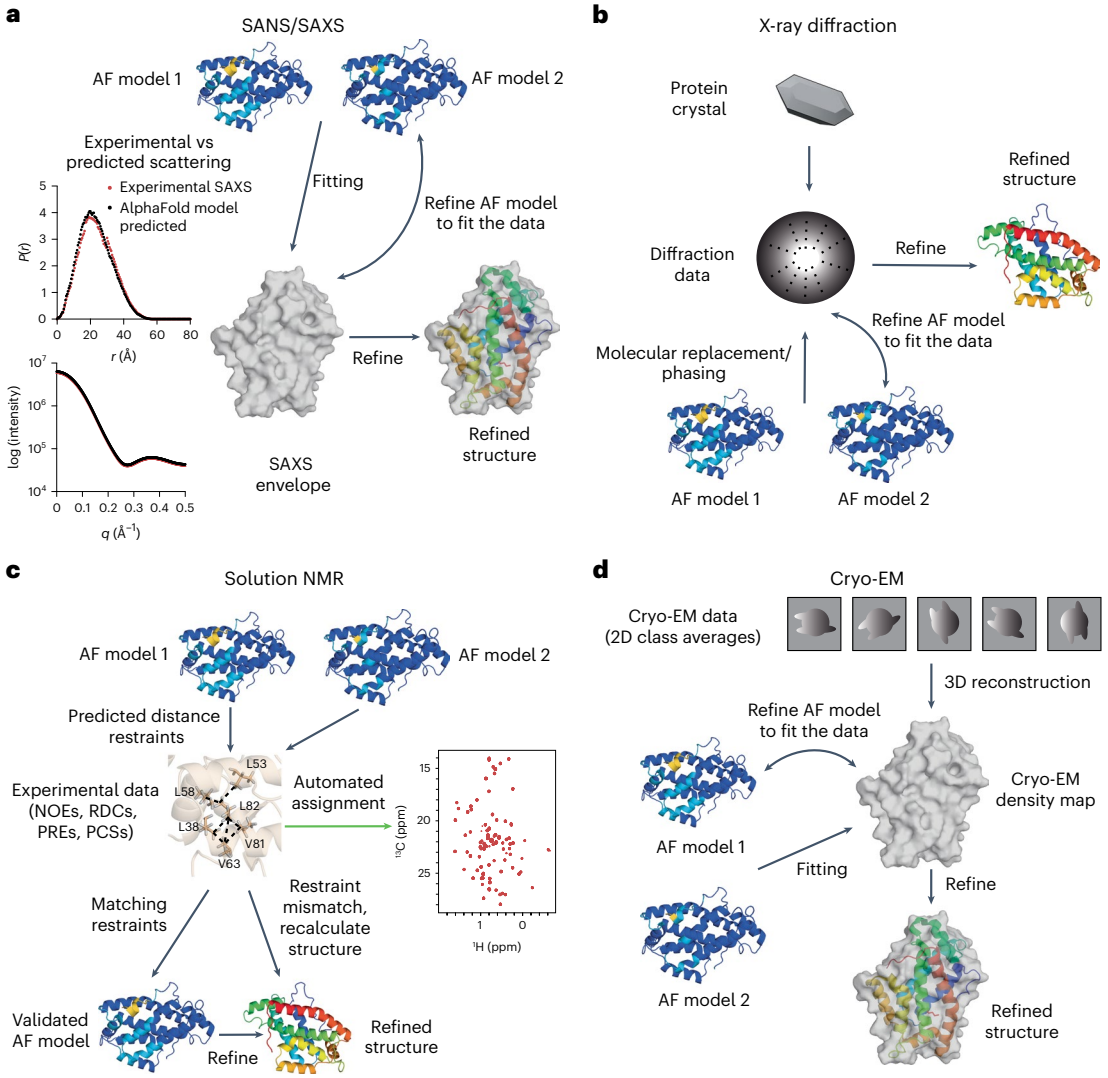


**Fig. 3 | Integration of AlphaFold models with experimental data. a**, Schematic of using AF2 models together with either SAXS or small-angle neutron scattering (SANS) data. In this example, AF2 models are compared against SAXS data in the form of the pair distribution function $P(r)$ and log ($I_q$) versus $q$ graphs. The SAXS envelope is fit together with AF2 models in an iterative fashion and refined to generate the final structure. AF, AlphaFold. **b**, Schematic of using AF2 models together with X-ray diffraction data. In the absence of an experimental template structure, AF2 models are iteratively used during the molecular replacement and/or phasing stages to process and fit the diffraction data in an iterative fashion. When the proper solution is found, the model is refined to generate the final structure. **c**, Schematic of using AF2 models together with solution NMR data. In one pathway, AF2 models are used together with experimental distance restraints (in the form of either NOEs, RDCs, paramagnetic relaxation enhancements (PREs), and/or pseudocontact shifts (PCSs) toward automated NMR resonance assignment via the predicted structure (in this case, a two-dimensional $^{1}$H-$^{13}$C[methyl] heteronuclear multiple quantum coherence spectra). In another pathway, predicted distances in the AF2 models are compared to those obtained experimentally. If the restraints match, the AF2 model is validated and refined. If the experimental restraints do not match, the AF2 model can be refined or recalculated using those restraints. **d**, Schematic of using AF2 models together with cryo-EM data. Two-dimensional (2D) class averages obtained from cryo-EM experiments are reconstructed into 3D density maps. AF2 models are iteratively fitted into the cryo-EM density map and refined to generate the final structure. For **a**–**d**, all graphs or maps shown are conceptual (that is, not real data). For all theoretical examples shown, the human glycolipid transfer protein (PDB ID 1SWX) was used.

Another burgeoning area where AlphaFold models are integrated with experimental data is solution-state NMR[15,27,28,30,81–85]. A series of AF2 models can be compared with experimental NMR data in the form of distance- and conformation-sensitive structural restraints obtained from NOEs[81], RDCs[29,86], paramagnetic relaxation enhancements[87,88] and/or pseudocontact shifts[89] (Fig. 3c). If NMR-derived restraints match the AF2 model, the structure can be refined. Otherwise, the NMR-derived restraints can be used to recalculate the structure using the AF2 model as a template[28]. Moreover, in the absence of NMR resonance assignments, AF2 models can be used as structural templates toward automated assignment[90]. This is especially helpful for large biological assemblies where methyl side chain labeling affords an increase in signal and resolution[91,92]. Here, NMR assignments for methyl side chain groups can be obtained using only methyl–methyl NOEs obtained from 3D NMR experiments and the atomic coordinates of a structure (or AF2-predicted structure) as input with software such as MAUS, MAGIC and methylFLYA[91,93,94].

AlphaFold models have also been used extensively together with single-particle cryo-EM data[60,76,95,96] (Fig. 3d). Two-dimensional class averages generated from tens of thousands of particle images are used as the input for 3D classification and reconstruction. A series of AF2 models are fitted into the 3D cryo-EM density maps, and each is evaluated for goodness of fit[96] and can be refined to generate a final structure[97]. Similar to X-ray diffraction studies, implicit incorporation of AF2 models, which are iteratively rebuilt on the basis of cryo-EM data, enables swift and robust structure modeling relative to ab initio model

building[60,98]. Here, the resolution of the cryo-EM data is essential for accurate model building. Reports suggest that AF2 models should not be fit into cryo-EM maps with resolution greater than 6 Å[98,99].

## Beyond static snapshots: ensembles and conformational landscapes

Native conformations of proteins are often described as time-averaged ensembles of conformations with Boltzmann-type distributions, especially IDPs and proteins with IDRs[100,101]. Apart from IDPs, well-folded globular proteins, such as G-protein-coupled receptors and kinases, also sample a wide range of conformations to carry out their biological function[102,103]. The standard implementation of AlphaFold performs well only in detecting a single structural snapshot (the 'ground state' structure), likely due to the lack of a large set of redundant protein conformers in the training set[16,32]. Thus, several groups have worked to extend AlphaFold to include predictions of structural ensembles and excited state structures[102,104]. Initial efforts to enhance sampling of different conformations have involved altering the number of sequences used to generate shallower MSA representations, masking coevolutionary information provided by MSAs and splitting conflicting coevolutionary signals by clustering MSAs[46,47,105]. Another approach is to enable dropout layers in the neural network, which are usually commonly used only during neural network training[8,106,107]. These approaches have shown great promise in increasing the ability of AF2 to predict alternative conformations, although benchmarking has been limited by the small number of structures in the PDB solved with multiple conformations. The use of experimental data, especially SAXS, NMR and cryo-EM, has also been described to guide modeling of ensembles and alternative conformations[71,82,108].

## Conclusions and outlook

AlphaFold and other machine learning-based structure-prediction software represents a giant leap forward in our understanding of protein function and structures. However, they are not yet 'one size fits all' solutions to the protein structure-prediction problem. Current implementations of AF2 can provide highly accurate working models for most rigid, well-folded globular proteins but may have issues predicting other classes of proteins. However, as suggested by recent work, we expect incredible progress in other classes of proteins in the coming years[33,37,47,51]. Machine learning approaches are also expected to be applied toward structure prediction of biomolecules, including nucleic acids[109], carbohydrates[110] and lipids. The case studies highlighted here reveal why caution must be taken in naive interpretation of AF2 models, even for cases with reasonable pLDDT and PAE confidence metrics (Fig. 2a–l). We expect that future studies will enable further refinement of error categorizations by teasing out fine details of cases with good evaluation metrics that do not match experimental results. We also expect to see increased integration of experimental data with AF2 predictions. Several of the studies mentioned here also show that simple modifications in the AF2 workflow can further extend its accuracy and applications into new horizons. On average, >10,000 protein structures are released in the PDB per year (https://www.rcsb.org/stats/growth/growth-protein). AF2 will continue to be evaluated against new experimental structures to further identify areas for improvement. Even in the face of an impressive display of accuracy, AF2 is still best used to complement and extend interpretation of experimental data at both structural and functional levels.

## Data availability

PyMOL sessions containing comparisons of AlphaFold models (extracted from the literature or the AlphaFold database) compared with experimental structures together with python script used to color code structures based on pLDDT values are freely available at https://github.com/mcshanlab/AlphaFold_Models_Agarwal_McShan.

## References


1. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
2. Bertoline, L. M. F., Lima, A. N., Krieger, J. E. & Teixeira, S. K. Before and after AlphaFold2: an overview of protein structure prediction. *Front. Bioinform.* **3**, 1120370 (2023).
3. Perrakis, A. & Sixma, T. K. AI revolutions in biology. *EMBO Rep.* **22**, e54046 (2021).
4. Bouatta, N., Sorger, P. & AlQuraishi, M. Protein structure prediction by AlphaFold2: are attention and symmetries all you need? *Acta Crystallogr. D* **77**, 982–991 (2021).
5. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
6. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
7. Varadi, M. et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52**, D368–D375 (2024).
8. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
9. Ahdritz, G. et al. OpenFold: retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *Nat. Methods* https://doi.org/10.1038/s41592-024-02272-z (2024).
10. Chen, S.-J. et al. Protein folds vs. protein folding: differing questions, different challenges. *Proc. Natl Acad. Sci. USA* **120**, e2214423119 (2023).
11. Skolnick, J., Gao, M., Zhou, H. & Singh, S. AlphaFold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. *J. Chem. Inf. Model.* **61**, 4827–4831 (2021).
12. Outeiral, C., Nissley, D. A. & Deane, C. M. Current structure predictors are not learning the physics of protein folding. *Bioinformatics* **38**, 1881–1887 (2022).
13. Alford, R. F. et al. The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
14. Roney, J. P. & Ovchinnikov, S. State-of-the-art estimation of protein model accuracy using AlphaFold. *Phys. Rev. Lett.* **129**, 238101 (2022).
15. Laurents, D. V. AlphaFold 2 and NMR spectroscopy: partners to understand protein structure, dynamics and function. *Front. Mol. Biosci.* **9**, 906437 (2022).
16. Chakravarty, D. & Porter, L. L. AlphaFold2 fails to predict protein fold switching. *Protein Sci.* **31**, e4353 (2022).
17. Akdel, M. et al. A structural biology community assessment of AlphaFold2 applications. *Nat. Struct. Mol. Biol.* **29**, 1056–1067 (2022).
18. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
19. Lin, Z. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
20. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.04.463034 (2022).
21. Gao, M., Nakajima An, D., Parks, J. M. & Skolnick, J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat. Commun.* **13**, 1744 (2022).
22. 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).

23. Oeffner, R. D. et al. Putting AlphaFold models to work with phenix. process_predicted_model and ISOLDE. *Acta Crystallogr. D* **78**, 1303–1314 (2022).
24. Tsaban, T. et al. Harnessing protein folding neural networks for peptide–protein docking. *Nat. Commun.* **13**, 176 (2022).
25. McDonald, E. F., Jones, T., Plate, L., Meiler, J. & Gulsevin, A. Benchmarking AlphaFold2 on peptide structure prediction. *Structure* **31**, 111–119 (2023).
26. Mikhaylov, V. et al. Accurate modeling of peptide–MHC structures with AlphaFold. *Structure* **32**, 228–241 (2024).
27. Alderson, T. R., Pritišanac, I., Kolarić, Đ., Moses, A. M. & Forman-Kay, J. D. Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proc. Natl Acad. Sci. USA* **120**, e2304302120 (2023).
28. Fowler, N. J. & Williamson, M. P. The accuracy of protein structures in solution determined by AlphaFold and NMR. *Structure* **30**, 925–933 (2022).
29. Zweckstetter, M. NMR hawk-eyed view of AlphaFold2 structures. *Protein Sci.* **30**, 2333–2337 (2021).
30. Tejero, R., Huang, Y. J., Ramelot, T. A. & Montelione, G. T. AlphaFold models of small proteins rival the accuracy of solution NMR Structures. *Front. Mol. Biosci.* **9**, 877000 (2022).
31. Thornton, J. M., Laskowski, R. A. & Borkakoti, N. AlphaFold heralds a data-driven revolution in biology and medicine. *Nat. Med.* **27**, 1666–1669 (2021).
32. Saldaño, T. et al. Impact of protein conformational diversity on AlphaFold predictions. *Bioinformatics* **38**, 2742–2748 (2022).
33. Hekkelman, M. L., de Vries, I., Joosten, R. P. & Perrakis, A. AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nat. Methods* **20**, 205–213 (2023).
34. Karelina, M., Noh, J. J. & Dror, R. O. How accurately can one predict drug binding modes using AlphaFold models? *eLife* **12**, RP89386 (2023).
35. Diwan, G. D., Gonzalez-Sanchez, J. C., Apic, G. & Russell, R. B. Next generation protein structure predictions and genetic variant interpretation. *J. Mol. Biol.* **433**, 167180 (2021).
36. David, A., Islam, S., Tankhilevich, E. & Sternberg, M. J. E. The AlphaFold database of protein structures: a biologist's guide. *J. Mol. Biol.* **434**, 167336 (2022).
37. Jambrich, M. A., Tusnady, G. E. & Dobson, L. How AlphaFold2 shaped the structural coverage of the human transmembrane proteome. *Sci. Rep.* **13**, 20283 (2023).
38. Hegedűs, T., Geisler, M., Lukács, G. L. & Farkas, B. Ins and outs of AlphaFold2 transmembrane protein structure predictions. *Cell. Mol. Life Sci.* **79**, 73 (2022).
39. Topitsch, A., Schwede, T. & Pereira, J. Outer membrane β-barrel structure prediction through the lens of AlphaFold2. *Proteins* **92**, 3–14 (2024).
40. Azzaz, F., Yahi, N., Chahinian, H. & Fantini, J. The epigenetic dimension of protein structure is an intrinsic weakness of the AlphaFold program. *Biomolecules* **12**, 1527 (2022).
41. Kishi, K. E. et al. Structural basis for channel conduction in the pump-like channelrhodopsin ChRmine. *Cell* **185**, 672–689 (2022).
42. Ruff, K. M. & Pappu, R. V. AlphaFold and implications for intrinsically disordered proteins. *J. Mol. Biol.* **433**, 167208 (2021).
43. Wilson, C. J., Choy, W.-Y. & Karttunen, M. AlphaFold2: a role for disordered protein/region prediction? *Int. J. Mol. Sci.* **23**, 4591 (2022).
44. Piovesan, D., Monzon, A. M. & Tosatto, S. C. E. Intrinsic protein disorder and conditional folding in AlphaFoldDB. *Protein Sci.* **31**, e4466 (2022).
45. Lane, T. J. Protein structure prediction has reached the single-structure frontier. *Nat. Methods* **20**, 170–173 (2023).
46. Del Alamo, D., Sala, D., Mchaourab, H. S. & Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* **11**, e75751 (2022).
47. Wayment-Steele, H. K. et al. Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **625**, 832–839 (2024).
48. Zhao, B., Ghadermarzi, S. & Kurgan, L. Comparative evaluation of AlphaFold2 and disorder predictors for prediction of intrinsic disorder, disorder content and fully disordered proteins. *Comput. Struct. Biotechnol. J.* **21**, 3248–3258 (2023).
49. Buel, G. R. & Walters, K. J. Can AlphaFold2 predict the impact of missense mutations on structure? *Nat. Struct. Mol. Biol.* **29**, 1–2 (2022).
50. Pak, M. A. et al. Using AlphaFold to predict the impact of single mutations on protein stability and function. *PLoS ONE* **18**, e0282689 (2023).
51. Cheng, J. et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* **381**, eadg7492 (2023).
52. Moffat, L., Greener, J. G. & Jones, D. T. Using AlphaFold for rapid and accurate fixed backbone protein design. Preprint at *bioRxiv* https://doi.org/10.1101/2021.08.24.457549 (2021).
53. Goverde, C. A., Wolf, B., Khakzad, H., Rosset, S. & Correia, B. E. De novo protein design by inversion of the AlphaFold structure prediction network. *Protein Sci.* **32**, e4653 (2023).
54. Yin, R., Feng, B. Y., Varshney, A. & Pierce, B. G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci.* **31**, e4379 (2022).
55. Yin, R. & Pierce, B. G. Evaluation of AlphaFold antibody–antigen modeling with implications for improving predictive accuracy. *Protein Sci.* **33**, e4865 (2024).
56. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein–protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022).
57. Jeppesen, M. & André, I. Accurate prediction of protein assembly structure by combining AlphaFold and symmetrical docking. *Nat. Commun.* **14**, 8283 (2023).
58. Pinheiro, F., Santos, J. & Ventura, S. AlphaFold and the amyloid landscape. *J. Mol. Biol.* **433**, 167059 (2021).
59. Binder, J. L. et al. AlphaFold Illuminates half of the dark human proteins. *Curr. Opin. Struct. Biol.* **74**, 102372 (2022).
60. Terwilliger, T. C. et al. Improved AlphaFold modeling with implicit experimental information. *Nat. Methods* **19**, 1376–1382 (2022).
61. Terwilliger, T. C. et al. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat. Methods* **21**, 110–116 (2024).
62. McCafferty, C. L., Pennington, E. L., Papoulas, O., Taylor, D. W. & Marcotte, E. M. Does AlphaFold2 model proteins' intracellular conformations? An experimental test using cross-linking mass spectrometry of endogenous ciliary proteins. *Commun. Biol.* **6**, 421 (2023).
63. Motmaen, A. et al. Peptide-binding specificity prediction using fine-tuned protein structure prediction networks. *Proc. Natl Acad. Sci. USA* **120**, e2216697120 (2023).
64. Jussupow, A. & Kaila, V. R. I. Effective molecular dynamics from neural network-based structure prediction models. *J. Chem. Theory Comput.* **19**, 1965–1975 (2023).
65. Guo, H.-B. et al. AlphaFold2 models indicate that protein sequence determines both structure and dynamics. *Sci. Rep.* **12**, 10696 (2022).
66. Carugo, O. pLDDT values in AlphaFold2 protein models are unrelated to globular protein local flexibility. *Crystals* **13**, 1560 (2023).
67. Zhu, W., Shenoy, A., Kundrotas, P. & Elofsson, A. Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes. *Bioinformatics* **39**, btad424 (2023).

68. Fontana, P. et al. Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold. *Science* **376**, eabm9326 (2022).
69. Terwilliger, T. C. et al. Accelerating crystal structure determination with iterative AlphaFold prediction. *Acta Crystallogr. D* **79**, 234–244 (2023).
70. Blanc, M. et al. Designed ankyrin repeat proteins provide insights into the structure and function of CagI and are potent inhibitors of CagA translocation by the *Helicobacter pylori* type IV secretion system. *PLoS Pathog.* **19**, e1011368 (2023).
71. Brookes, E., Rocco, M., Vachette, P. & Trewhella, J. AlphaFold-predicted protein structures and small-angle X-ray scattering: insights from an extended examination of selected data in the Small-Angle Scattering Biological Data Bank. *J. Appl. Crystallogr.* **56**, 910–926 (2023).
72. Brookes, E. & Rocco, M. A database of calculated solution parameters for the AlphaFold predicted protein structures. *Sci. Rep.* **12**, 7349 (2022).
73. Chinnam, N. B. et al. Combining small angle X-ray scattering (SAXS) with protein structure predictions to characterize conformations in solution. *Methods Enzymol.* **678**, 351–376 (2023).
74. Da Vela, S. & Svergun, D. I. Methods, development and applications of small-angle X-ray scattering to characterize biological macromolecules in solution. *Curr. Res. Struct. Biol.* **2**, 164–170 (2020).
75. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477–481 (2013).
76. Kryshtafovych, A. et al. Computational models in the service of X-ray and cryo-EM structure determination. *Proteins* **89**, 1633–1646 (2021).
77. Chai, L. et al. AlphaFold protein structure database for sequence-independent molecular replacement. *Crystals* **11**, 1227 (2021).
78. McCoy, A. J., Sammito, M. D. & Read, R. J. Implications of AlphaFold2 for crystallographic phasing by molecular replacement. *Acta Crystallogr. D* **78**, 1–13 (2022).
79. Barbarin-Bocahu, I. & Graille, M. The X-ray crystallography phase problem solved thanks to AlphaFold and RoseTTAFold models: a case-study report. *Acta Crystallogr. D* **78**, 517–531 (2022).
80. Abergel, C. Molecular replacement: tricks and treats. *Acta Crystallogr. D* **69**, 2167–2173 (2013).
81. Chiliveri, S. C. et al. Experimental NOE, chemical shift, and proline isomerization data provide detailed insights into amelotin oligomerization. *J. Am. Chem. Soc.* **145**, 18063–18074 (2023).
82. Abdollahi, H., Prestegard, J. H. & Valafar, H. Computational modeling multiple conformational states of proteins with residual dipolar coupling data. *Curr. Opin. Struct. Biol.* **82**, 102655 (2023).
83. Sedinkin, S. L., Burns, D., Shukla, D., Potoyan, D. A. & Venditti, V. Solution structure ensembles of the open and closed forms of the ~130 kDa enzyme I via AlphaFold modeling, coarse grained simulations, and NMR. *J. Am. Chem. Soc.* **145**, 13347–13356 (2023).
84. Li, E. H. et al. Blind assessment of monomeric AlphaFold2 protein structure models with experimental NMR data. *J. Magn. Reson.* **352**, 107481 (2023).
85. Ma, P., Li, D.-W. & Brüschweiler, R. Predicting protein flexibility with AlphaFold. *Proteins* **91**, 847–855 (2023).
86. Robertson, A. J., Courtney, J. M., Shen, Y., Ying, J. & Bax, A. Concordance of X-ray and AlphaFold2 models of SARS-CoV-2 main protease with residual dipolar couplings measured in solution. *J. Am. Chem. Soc.* **143**, 19306–19310 (2021).
87. Lenard, A. J., Mulder, F. A. A. & Madl, T. Solvent paramagnetic relaxation enhancement as a versatile method for studying structure and dynamics of biomolecular systems. *Prog. Nucl. Magn. Reson. Spectrosc.* **132–133**, 113–139 (2022).
88. Koehler Leman, J. & Künze, G. Recent advances in NMR protein structure prediction with Rosetta. *Int. J. Mol. Sci.* **24**, 7835 (2023).
89. Zhu, W., Yang, D. T. & Gronenborn, A. M. Ligand-capped cobalt(II) multiplies the value of the double-histidine motif for PCS NMR studies. *J. Am. Chem. Soc.* **145**, 4564–4569 (2023).
90. Klukowski, P., Riek, R. & Güntert, P. Time-optimized protein NMR assignment with an integrative deep learning approach using AlphaFold and chemical shift prediction. *Sci. Adv.* **9**, eadi9323 (2023).
91. McShan, A. C. Utility of methyl side chain probes for solution NMR studies of large proteins. *J. Magn. Reson. Open* **14–15**, 100087 (2023).
92. Ruschak, A. M. & Kay, L. E. Methyl groups as probes of supra-molecular structure, dynamics and function. *J. Biomol. NMR* **46**, 75–87 (2009).
93. Pritišanac, I., Würz, J. M., Alderson, T. R. & Güntert, P. Automatic structure-based NMR methyl resonance assignment in large proteins. *Nat. Commun.* **10**, 4922 (2019).
94. Clay, M. C., Saleh, T., Kamatham, S., Rossi, P. & Kalodimos, C. G. Progress toward automated methyl assignments for methyl-TROSY applications. *Structure* **30**, 69–79 (2022).
95. Giri, N., Roy, R. S. & Cheng, J. Deep learning for reconstructing protein structures from cryo-EM density maps: recent advances and future directions. *Curr. Opin. Struct. Biol.* **79**, 102536 (2023).
96. Hryc, C. F. & Baker, M. L. AlphaFold2 and cryoEM: revisiting cryoEM modeling in near-atomic resolution density maps. *iScience* **25**, 104496 (2022).
97. Reggiano, G., Lugmayr, W., Farrell, D., Marlovits, T. C. & DiMaio, F. Residue-level error detection in cryo-electron microscopy models. *Structure* **31**, 860–869 (2023).
98. Dai, X., Wu, L., Yoo, S. & Liu, Q. Integrating AlphaFold and deep learning for atomistic interpretation of cryo-EM maps. *Brief. Bioinform.* **24**, bbad405 (2023).
99. Alshammari, M., He, J. & Wriggers, W. Appraisal of AlphaFold2-predicted models in cryo-EM map interpretation. *Microsc. Microanal.* **29**, 977–978 (2023).
100. Lindorff-Larsen, K. & Kragelund, B. B. On the potential of machine learning to examine the relationship between sequence, structure, dynamics and function of intrinsically disordered proteins. *J. Mol. Biol.* **433**, 167196 (2021).
101. Wei, G., Xi, W., Nussinov, R. & Ma, B. Protein ensembles: how does nature harness thermodynamic fluctuations for life? The diverse functional roles of conformational ensembles in the cell. *Chem. Rev.* **116**, 6516–6551 (2016).
102. Sala, D., Hildebrand, P. W. & Meiler, J. Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. *Front. Mol. Biosci.* **10**, 1121962 (2023).
103. Heo, L. & Feig, M. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins* **90**, 1873–1885 (2022).
104. Sala, D., Engelberger, F., Mchaourab, H. S. & Meiler, J. Modeling conformational states of proteins with AlphaFold. *Curr. Opin. Struct. Biol.* **81**, 102645 (2023).
105. Stein, R. A. & Mchaourab, H. S. SPEACH_AF: sampling protein ensembles and conformational heterogeneity with AlphaFold2. *PLoS Comput. Biol.* **18**, e1010483 (2022).
106. Wallner, B. AFsample: improving multimer prediction with AlphaFold using massive sampling. *Bioinformatics* **39**, btad573 (2023).
107. Johansson-Åkhe, I. & Wallner, B. Improving peptide–protein docking with AlphaFold-Multimer using forced sampling. *Front. Bioinform.* **2**, 959160 (2022).

108. Ramelot, T. A., Tejero, R. & Montelione, G. T. Representing structures of the multiple conformational states of proteins. *Curr. Opin. Struct. Biol.* **83**, 102703 (2023).
109. Townshend, R. J. L. et al. Geometric deep learning of RNA structure. *Science* **373**, 1047–1051 (2021).
110. Bojar, D. & Lisacek, F. Glycoinformatics in the artificial intelligence era. *Chem. Rev.* **122**, 15971–15988 (2022).

## Acknowledgements

A.C.M. acknowledges start-up funds from the Georgia Institute of Technology. V.A. acknowledges support from the National Science Foundation (CHE-2238650) and the National Institutes of Health (R35GM142882).

## Author contributions

A.C.M. and V.A. conceived, wrote and edited the manuscript. A.C.M. generated figures and analyzed models.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence and requests for materials** should be addressed to Vinayak Agarwal or Andrew C. McShan.

**Peer review information** *Nature Chemical Biology* thanks Paul Adams and the other, anonymous reviewer(s) for their contribution to the peer review of this work.