Article

Improved protein structure prediction using potentials from deep learning

https://doi.org/10.1038/s41586-019-1923-7

Received: 2 April 2019

Accepted: 10 December 2019

Published online: 15 January 2020

Andrew W. Senior[1,4]*, Richard Evans[1,4], John Jumper[1,4], James Kirkpatrick[1,4], Laurent Sifre[1,4], Tim Green[1], Chongli Qin[1], Augustin Žídek[1], Alexander W. R. Nelson[1], Alex Bridgland[1], Hugo Penedones[1], Stig Petersen[1], Karen Simonyan[1], Steve Crossan[1], Pushmeet Kohli[1], David T. Jones[2,3], David Silver[1], Koray Kavukcuoglu[1] & Demis Hassabis[1]

Protein structure prediction can be used to determine the three-dimensional shape of a protein from its amino acid sequence[1]. This problem is of fundamental importance as the structure of a protein largely determines its function[2]; however, protein structures can be difficult to determine experimentally. Considerable progress has recently been made by leveraging genetic information. It is possible to infer which amino acid residues are in contact by analysing covariation in homologous sequences, which aids in the prediction of protein structures[3]. Here we show that we can train a neural network to make accurate predictions of the distances between pairs of residues, which convey more information about the structure than contact predictions. Using this information, we construct a potential of mean force[4] that can accurately describe the shape of a protein. We find that the resulting potential can be optimized by a simple gradient descent algorithm to generate structures without complex sampling procedures. The resulting system, named AlphaFold, achieves high accuracy, even for sequences with fewer homologous sequences. In the recent Critical Assessment of Protein Structure Prediction[5] (CASP13)—a blind assessment of the state of the field—AlphaFold created high-accuracy structures (with template modelling (TM) scores[6] of 0.7 or higher) for 24 out of 43 free modelling domains, whereas the next best method, which used sampling and contact information, achieved such accuracy for only 14 out of 43 domains. AlphaFold represents a considerable advance in protein-structure prediction. We expect this increased accuracy to enable insights into the function and malfunction of proteins, especially in cases for which no structures for homologous proteins have been experimentally determined[7].

Proteins are at the core of most biological processes. As the function of a protein is dependent on its structure, understanding protein structures has been a grand challenge in biology for decades. Although several experimental structure determination techniques have been developed and improved in accuracy, they remain difficult and time-consuming[2]. As a result, decades of theoretical work has attempted to predict protein structures from amino acid sequences.

CASP[5] is a biennial blind protein structure prediction assessment run by the structure prediction community to benchmark progress in accuracy. In 2018, AlphaFold joined 97 groups from around the world in entering CASP13[8]. Each group submitted up to 5 structure predictions for each of 84 protein sequences for which experimentally determined structures were sequestered. Assessors divided the proteins into 104 domains for scoring and classified each as being amenable to template-based modelling (TBM, in which a protein with a similar sequence has a known structure, and that homologous structure is modified in accordance with the sequence differences) or requiring free modelling (FM, in cases in which no homologous structure is available), with an intermediate (FM/TBM) category. Figure 1a shows that AlphaFold predicts more FM domains with high accuracy than any other system, particularly in the 0.6–0.7 TM-score range. The TM score—ranging between 0 and 1—measures the degree of match of the overall (backbone) shape of a proposed structure to a native structure. The assessors ranked the 98 participating groups by the summed, capped *z*-scores of the structures, separated according to category. AlphaFold achieved a summed *z*-score of 52.8 in the FM category (best-of-five) compared with 36.6 for the next closest group (322). Combining FM and TBM/FM categories, AlphaFold scored 68.3 compared with 48.2. AlphaFold is able to predict previously unknown folds to high accuracy (Fig. 1b). Despite using only FM techniques and not using templates, AlphaFold also scored well in the TBM category according to the assessors' formula 0-capped *z*-score, ranking fourth for the top-one model or first for the best-of-five models. Much of the accuracy of AlphaFold is due to the accuracy of the distance predictions, which is evident from the high precision of the corresponding contact predictions (Fig. 1c and Extended Data Fig. 2a).

[1]DeepMind, London, UK. [2]The Francis Crick Institute, London, UK. [3]University College London, London, UK. [4]These authors contributed equally: Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre. *e-mail: andrewsenior@google.com

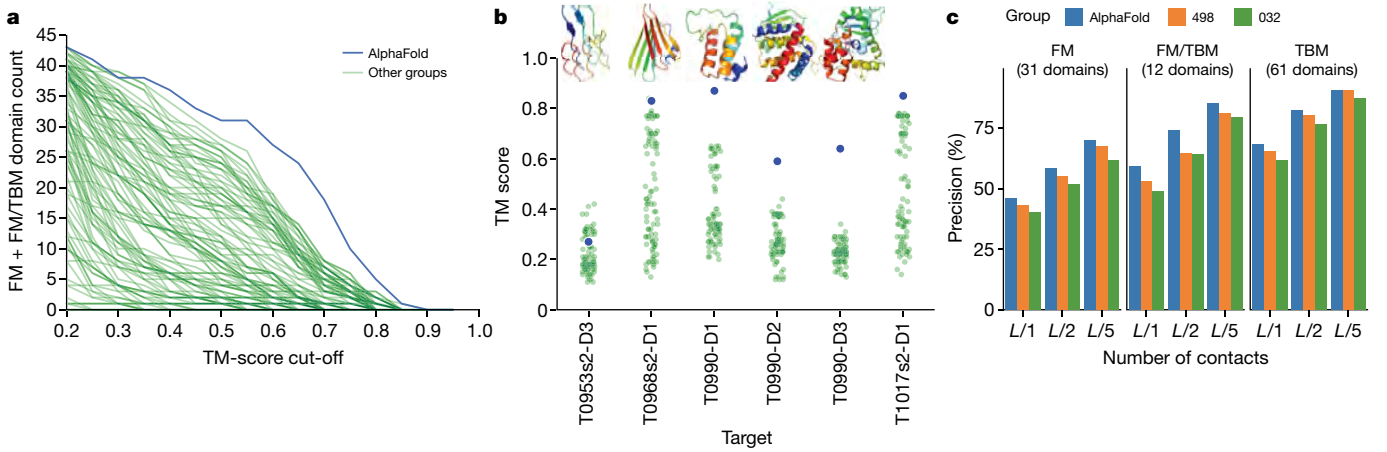


Fig. 1 | The performance of AlphaFold in the CASP13 assessment. a, Number of FM (FM + FM/TBM) domains predicted for a given TM-score threshold for AlphaFold and the other 97 groups. **b**, For the six new folds identified by the CASP13 assessors, the TM score of AlphaFold was compared with the other groups, together with the native structures. The structure of T1017s2-D1 is not available for publication. **c**, Precisions for long-range contact prediction in CASP13 for the most probable L, $L/2$ or $L/5$ contacts, where L is the length of the domain. The distance distributions used by AlphaFold in CASP13, thresholded to contact predictions, are compared with the submissions by the two best-ranked contact prediction methods in CASP13: 498 (RaptorX-Contact[26]) and 032 (TripletRes[32]) on 'all groups' targets, with updated domain definitions for T0953s2.

The most-successful FM approaches thus far[9–11] have relied on fragment assembly. In these approaches, a structure is created through a stochastic sampling process—such as simulated annealing[12]—that minimizes a statistical potential that is derived from summary statistics extracted from structures in the Protein Data Bank (PDB)[13]. In fragment assembly, a structure hypothesis is repeatedly modified, typically by changing the shape of a short section while retaining changes that lower the potential, ultimately leading to low potential structures. Simulated annealing requires many thousands of such moves and must be repeated many times to have good coverage of low-potential structures.

In recent years, the accuracy of structure predictions has improved through the use of evolutionary covariation data[14] that are found in sets of related sequences. Sequences that are similar to the target sequence are found by searching large datasets of protein sequences derived from DNA sequencing and aligned to the target sequence to generate a multiple sequence alignment (MSA). Correlated changes in the positions of two amino acid residues across the sequences of the MSA can be used to infer which residues might be in contact. Contacts are typically defined to occur when the β-carbon atoms of 2 residues are within 8 Å of one another. Several methods[15–18], including neural networks[19–22], have been used to predict the probability that a pair of residues is in contact based on features computed from MSAs. Contact predictions are incorporated in structure predictions by modifying the statistical potential to guide the folding process to structures that satisfy more of the predicted contacts[11,23]. Other studies[24,25] have used predictions of the distance between residues, particularly for distance geometry approaches[26–28]. Neural network distance predictions without covariation features were used to make the evolutionary pairwise distance-dependent statistical potential[25], which was used to rank structure hypotheses. In addition, the QUARK pipeline[11] used a template-based distance-profile restraint for TBM.

In this study, we present a deep-learning approach to protein structure prediction, the stages of which are illustrated in Fig. 2a. We show that it is possible to construct a learned, protein-specific potential by training a neural network (Fig. 2b) to make accurate predictions about the structure of the protein given its sequence, and to predict the structure itself accurately by minimizing the potential by gradient descent (Fig. 2c). The neural network predictions include backbone torsion angles and pairwise distances between residues. Distance predictions provide more specific information about the structure than contact predictions and provide a richer training signal for the neural network. By jointly predicting many distances, the network can propagate distance information that respects covariation, local structure and residue identities of nearby residues. The predicted probability distributions can be combined to form a simple, principled protein-specific potential. We show that with gradient descent, it is simple to find a set of torsion angles that minimizes this protein-specific potential using only limited sampling. We also show that whole chains can be optimized simultaneously, avoiding the need to segment long proteins into hypothesized domains that are modelled independently as is common practice (see Methods).

The central component of AlphaFold is a convolutional neural network that is trained on PDB structures to predict the distances d_{ij} between the C_β atoms of pairs, ij, of residues of a protein. On the basis of a representation of the amino acid sequence, S, of a protein and features derived from the MSA(S) of that sequence, the network, which is similar in structure to those used for image-recognition tasks[29], predicts a discrete probability distribution $P(d_{ij}|S, \mathrm{MSA}(S))$ for every ij pair in any 64 × 64 region of the $L \times L$ distance matrix, as shown in Fig. 2b. The full set of distance distribution predictions constructed by combining such predictions that covers the entire distance map is termed a distogram (from distance histogram). Example distogram predictions for one CASP protein, T0955, are shown in Fig. 3c, d. The modes of the distribution (Fig. 3c) can be seen to closely match the true distances (Fig. 3b). Example distributions for all distances to one residue (residue 29) are shown in Fig. 3d. We found that the predictions of the distance correlate well with the true distance between residues (Fig. 3e). Furthermore, the network also models the uncertainty in its predictions (Fig. 3f). When the s.d. of the predicted distribution is low, the predictions are more accurate. This is also evident in Fig. 3d, in which more confident predictions of the distance distribution (higher peak and lower s.d. of the distribution) tend to be more accurate, with the true distance close to the peak. Broader, less-confidently predicted distributions still assign probability to the correct value even when it is not close to the peak. The high accuracy of the distance predictions and consequently the contact predictions (Fig. 1c) comes from a combination of factors in the design of the neural network and its training, data augmentation, feature representation, auxiliary losses, cropping and data curation (see Methods).

To generate structures that conform to the distance predictions, we constructed a smooth potential V_{distance} by fitting a spline to the negative log probabilities, and summing across all of the residue pairs

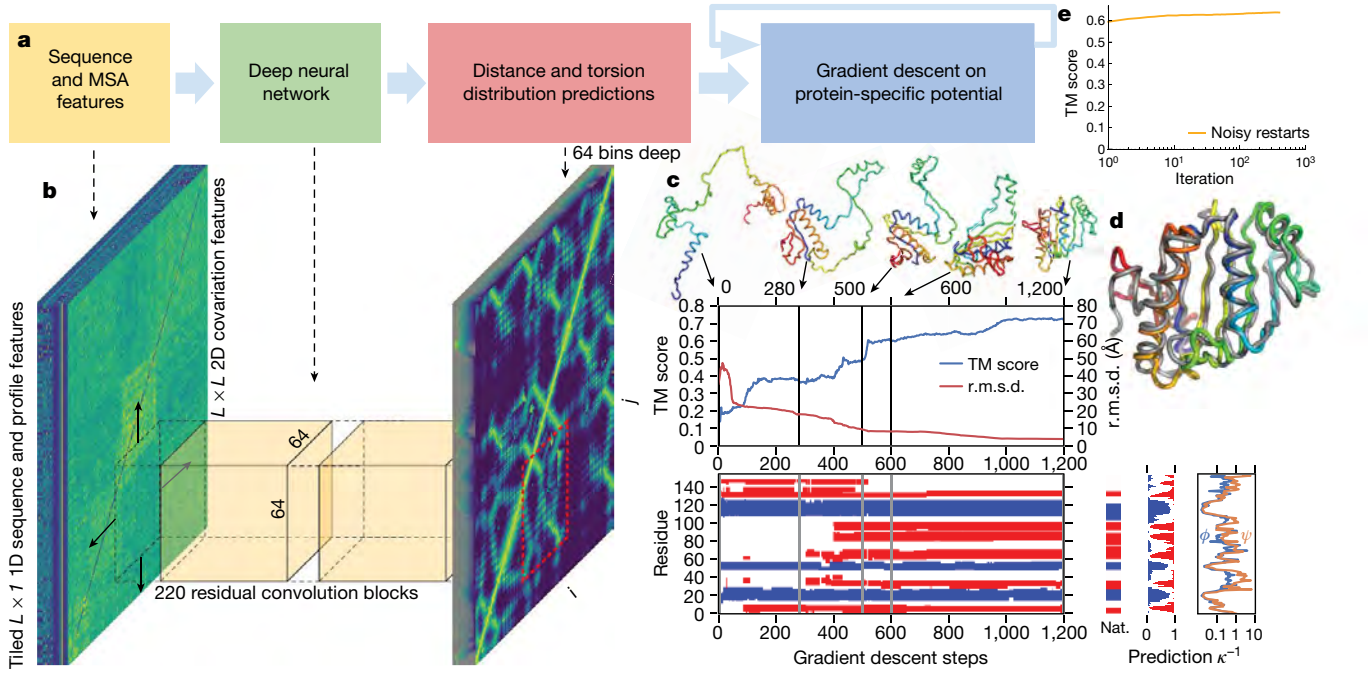


Fig. 2 | The folding process illustrated for CASP13 target T0986s2. CASP target T0986s2, $L = 155$, PDB: 6N9V. **a**, Steps of structure prediction. **b**, The neural network predicts the entire $L \times L$ distogram based on MSA features, accumulating separate predictions for 64 × 64-residue regions. **c**, One iteration of gradient descent (1,200 steps) is shown, with the TM score and root mean square deviation (r.m.s.d.) plotted against step number with five snapshots of the structure. The secondary structure (from SST[33]) is also shown (helix in blue, strand in red) along with the native secondary structure (Nat.), the secondary structure prediction probabilities of the network and the uncertainty in torsion angle predictions (as κ^{-1} of the von Mises distributions fitted to the predictions for φ and ψ). While each step of gradient descent greedily lowers the potential, large global conformation changes are effected, resulting in a well-packed chain. **d**, The final first submission overlaid on the native structure (in grey). **e**, The average (across the test set, $n = 377$) TM score of the lowest-potential structure against the number of repeats of gradient descent per target (log scale).

(see Methods). We parameterized protein structures by the backbone torsion angles (φ, ψ) of all residues and build a differentiable model of protein geometry $\mathbf{x} = G(\varphi, \psi)$ to compute the C_β coordinates, $\mathbf{x}_i$ for all residues i and thus the inter-residue distances, $d_{ij} = ||\mathbf{x}_i - \mathbf{x}_j||$, for each structure, and express V_{distance} as a function of φ and ψ. For a protein with L residues, this potential accumulates L^2 terms from marginal distribution predictions. To correct for the overrepresentation of the prior, we subtract a reference distribution[30] from the distance potential in the log domain. The reference distribution models the distance distributions $P(d_{ij}|\text{length})$ independent of the protein sequence and is computed by training a small version of the distance prediction neural network on the same structures, without sequence or MSA input features. A separate output head of the contact prediction network is trained to predict discrete probability distributions of backbone torsion angles $P(\varphi_i, \psi_i|S, \text{MSA}(S))$. After fitting a von Mises distribution, this is used to add a smooth torsion modelling term, V_{torsion}, to the potential. Finally, to prevent steric clashes, we add the $V_{\text{score2\_smooth}}$ score of Rosetta[9] to the potential, as this incorporates a van der Waals term. We used multiplicative weights for each of the three terms in the potential; however, no combination of weights noticeably outperformed equal weighting.

As all of the terms in the combined potential $V_{\text{total}}(\varphi, \psi)$ are differentiable functions of (φ, ψ), it can be optimized with respect to these variables by gradient descent. Here we use L-BFGS[31]. Structures are initialized by sampling torsion values from $P(\varphi_i, \psi_i|S, \text{MSA}(S))$. Figure 2c illustrates a single gradient descent trajectory that minimizes the potential, showing how this greedy optimization process leads to increasing accuracy and large-scale conformation changes. The secondary structure is partly set by the initialization from the predicted torsion angle distributions. The overall accuracy (TM score) improves quickly and after a few hundred steps of gradient descent the accuracy of the structure has converged to a local optimum of the potential.

We repeated the optimization from sampled initializations, leading to a pool of low-potential structures from which further structure initializations are sampled, with added backbone torsion noise ('noisy restarts'), leading to more structures to be added to the pool. After only a few hundred cycles, the optimization converges and the lowest potential structure is chosen as the best candidate structure. Figure 2e shows the progress in the accuracy of the best-scoring structures over multiple restarts of the gradient descent process, showing that after a few iterations the optimization has converged. Noisy restarts enable structures with a slightly higher TM score to be found than when continuing to sample from the predicted torsion distributions (average of 0.641 versus 0.636 on our test set, shown in Extended Data Fig. 4).

Figure 4a shows that the distogram accuracy (measured using the local distance difference test (lDDT_{12}) of the distogram; see Methods) correlates well with the TM score of the final realized structures. Figure 4b shows the effect of changing the construction of the potential. Removing the distance potential entirely gives a TM score of 0.266. Reducing the resolution of the distogram representation below six bins by averaging adjacent bins causes the TM score to degrade. Removing the torsion potential, reference correction or $V_{\text{score2\_smooth}}$ degrades the accuracy only slightly. A final 'relaxation' (side-chain packing interleaved with gradient descent) with Rosetta[9], using a combination of the Talaris2014 potential and a spline fit of our reference-corrected distance potential adds side-chain atom coordinates, and yields a small average improvement of 0.007 TM score.

We show that a carefully designed deep-learning system can provide accurate predictions of inter-residue distances and can be used to construct a protein-specific potential that represents the protein structure. Furthermore, we show that this potential can be optimized with gradient descent to achieve accurate structure predictions.

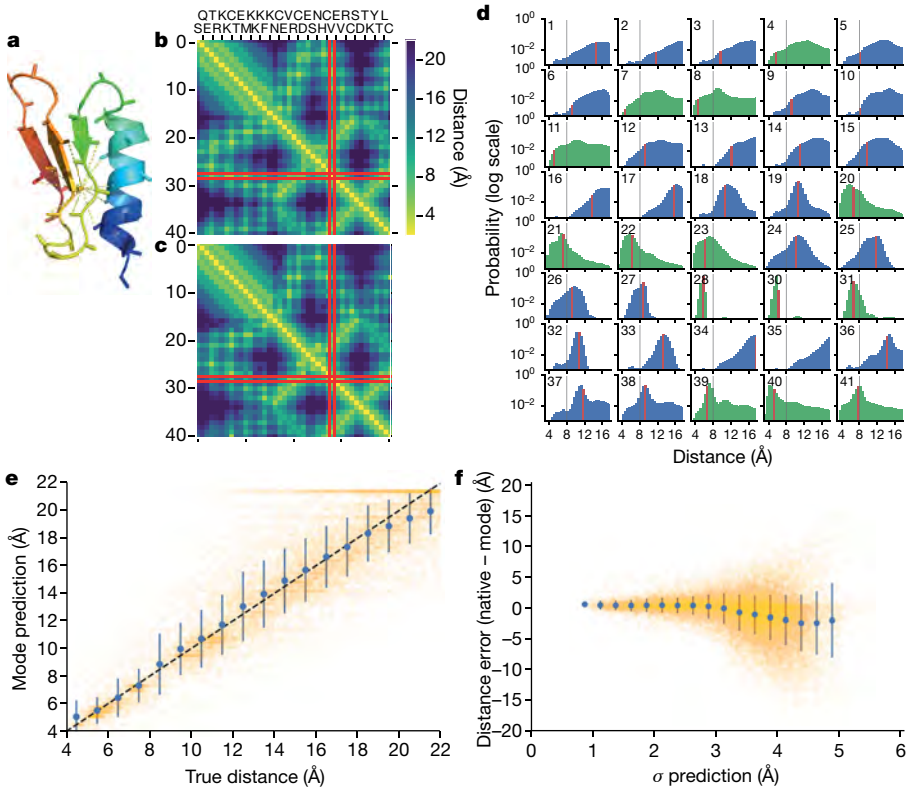


Fig. 3 | Predicted distance distributions compared with true distances. **a**–**d**, CASP target T0955, $L = 41$, PDB 5W9F. **a**, Native structure showing distances under 8 Å from the C_β of residue 29. **b**, **c**, Native inter-residue distances (**b**) and the mode of the distance predictions (**c**), highlighting residue 29. **d**, The predicted probability distributions for distances of residue 29 to all other residues. The bin corresponding to the native distance is highlighted in red, 8 Å is drawn in black. The distributions of the true contacts are plotted in green, non-contacts in blue. **e**, **f**, CASP target T0990, $L = 552$, PDB 6N9V. **e**, The mode of the predicted distance plotted against the true distance for all residue pairs with distances ≤22 Å, excluding distributions with s.d. > 3.5 Å ($n = 28{,}678$). Data are mean ± s.d. calculated for 1 Å bins. **f**, The error of the mode distance prediction versus the s.d. of the distance distributions, excluding pairs with native distances >22 Å ($n = 61{,}872$). Data are mean ± s.d. are shown for 0.25 Å bins. The true distance matrix and distogram for T0990 are shown in Extended Data Fig. 2b, c.

Whereas FM predictions only rarely approach the accuracy of experimental structures, the CASP13 assessment shows that the AlphaFold system achieves unprecedented FM accuracy and that this FM method can match the performance of template-modelling approaches without using templates and is starting to reach the accuracy needed to provide biological insights (see Methods). We hope that the methods we have

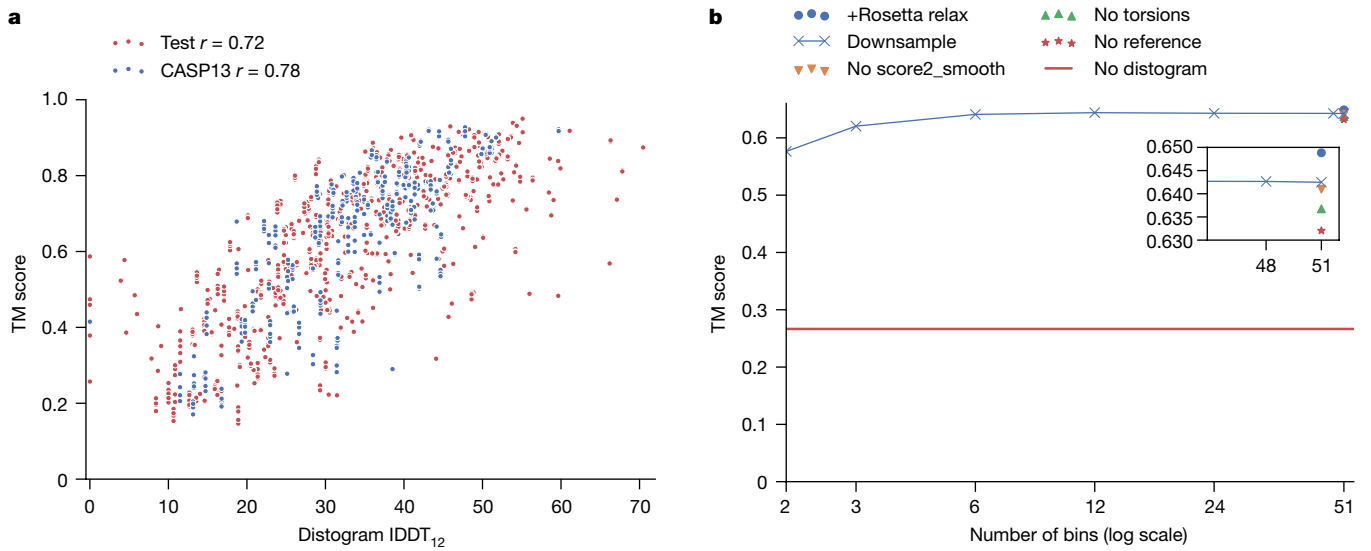


Fig. 4 | TM scores versus the accuracy of the distogram, and the dependency of the TM score on different components of the potential. **a**, TM score versus distogram $lDDT_{12}$ with Pearson's correlation coefficients, for both CASP13 ($n = 500$: 5 decoys for all domains, excluding T0999) and test ($n = 377$) datasets. **b**, Average TM score over the test set ($n = 377$) versus the number of histogram bins used when downsampling the distogram, compared with removing different components of the potential, or adding Rosetta relaxation.

described can be developed further and applied to benefit all areas of protein science with more accurate predictions for sequences of unknown structure.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1923-7.

1. Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. The protein folding problem. *Annu. Rev. Biophys.* **37**, 289–316 (2008).
2. Dill, K. A. & MacCallum, J. L. The protein-folding problem, *50 years on*. *Science* **338**, 1042–1046 (2012).
3. Schaarschmidt, J., Monastyrskyy, B., Kryshtafovych, A. & Bonvin, A. M. J. J. Assessment of contact predictions in CASP12: co-evolution and deep learning coming of age. *Proteins* **86**, 51–66 (2018).
4. Kirkwood, J. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **3**, 300–313 (1935).
5. Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. Critical assessment of methods of protein structure prediction (CASP)—Round XIII. *Proteins* **87**, 1011–1020 (2019).
6. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710 (2004).
7. Zhang, Y. Protein structure prediction: when is it useful? *Curr. Opin. Struct. Biol.* **19**, 145–155 (2009).
8. Senior, A. W. et al. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins* **87**, 1141–1148 (2019).
9. Das, R. & Baker, D. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77**, 363–382 (2008).
10. Jones, D. T. Predicting novel protein folds by using FRAGFOLD. *Proteins* **45**, 127–132 (2001).
11. Zhang, C., Mortuza, S. M., He, B., Wang, Y. & Zhang, Y. Template-based and free modeling of I-TASSER and QUARK pipelines using predicted contact maps in CASP12. *Proteins* **86**, 136–151 (2018).
12. Kirkpatrick, S., Gelatt, C. D. Jr & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
13. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
14. Altschuh, D., Lesk, A. M., Bloomer, A. C. & Klug, A. Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. *J. Mol. Biol.* **193**, 693–707 (1987).
15. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030 (2014).
16. Seemayer, S., Gruber, M. & Söding, J. CCMpred—fast and precise prediction of protein residue–residue contacts from correlated mutations. *Bioinformatics* **30**, 3128–3130 (2014).
17. Morcos, F. et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl Acad. Sci. USA* **108**, E1293–E1301 (2011).
18. Jones, D. T., Buchan, D. W., Cozzetto, D. & Pontil, M. PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics* **28**, 184–190 (2012).
19. Skwark, M. J., Raimondi, D., Michel, M. & Elofsson, A. Improved contact predictions using the recognition of protein like contact patterns. *PLOS Comput. Biol.* **10**, e1003889 (2014).
20. Jones, D. T., Singh, T., Kosciolek, T. & Tetchner, S. MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. *Bioinformatics* **31**, 999–1006 (2015).
21. Wang, S., Sun, S., Li, Z., Zhang, R. & Xu, J. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLOS Comput. Biol.* **13**, e1005324 (2017).
22. Jones, D. T. & Kandathil, S. M. High precision in protein contact prediction using fully convolutional neural networks and minimal sequence features. *Bioinformatics* **34**, 3308–3315 (2018).
23. Ovchinnikov, S. et al. Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Proteins* **84**, 67–75 (2016).
24. Aszódi, A. & Taylor, W. R. Estimating polypeptide α-carbon distances from multiple sequence alignments. *J. Math. Chem.* **17**, 167–184 (1995).
25. Zhao, F. & Xu, J. A position-specific distance-dependent statistical potential for protein structure and functional study. *Structure* **20**, 1118–1126 (2012).
26. Xu, J. & Wang, S. Analysis of distance-based protein structure prediction by deep learning in CASP13. *Proteins* **87**, 1069–1081 (2019).
27. Aszódi, A., Gradwell, M. J. & Taylor, W. R. Global fold determination from a small number of distance restraints. *J. Mol. Biol.* **251**, 308–326 (1995).
28. Kandathil, S. M., Greener, J. G. & Jones, D. T. Prediction of interresidue contacts with DeepMetaPSICOV in CASP13. *Proteins* **87**, 1092–1099 (2019).
29. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (2016).
30. Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225 (1997).
31. Liu, D. C. & Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Program.* **45**, 503–528 (1989).
32. Li, Y., Zhang, C., Bell, E. W., Yu, D.-J. & Zhang, Y. Ensembling multiple raw coevolutionary features with deep residual neural networks for contact-map prediction in CASP13. *Proteins* **87**, 1082–1091 (2019).
33. Konagurthu, A. S., Lesk, A. M. & Allison, L. Minimum message length inference of secondary structure from protein coordinate data. *Bioinformatics* **28**, i97–i105 (2012).

Methods

Extended Data Figure 1a shows the steps involved in MSA construction, feature extraction, distance prediction, potential construction and structure realization.

Tools

The following tools and dataset versions were used for the CASP system and for subsequent experiments: PDB 15 March 2018; CATH 16 March 2018; HHblits based on v.3.0-beta.3 (three iterations, $E=1\times10^{-3}$); HHpred web server; Uniclust30 2017-10; PSI-BLAST v.2.6.0 nr dataset (as of 15 December 2017) (three iterations, $E=1\times10^{-3}$); SST web server (March 2019); BioPython v.1.65; Rosetta v.3.5; PyMol 2.2.0 for structure visualization; TM-align 20160521.

Data

Our models are trained on structures extracted from the PDB[13]. We extract non-redundant domains by utilizing the CATH[34] 35% sequence similarity cluster representatives. This generated 31,247 domains, which were split into train and test sets (29,427 and 1,820 proteins, respectively), keeping all domains from the same homologous superfamily (H-level in the CATH classification) in the same partition. The CATH superfamilies of FM domains from CASP11 and CASP12 were also excluded from the training set. From the test set, we took—at random—a single domain per homologous superfamily to create the 377 domain subset used for the results presented here. We note that accuracies for this set are higher than for the CASP13 test domains.

CASP13 submission results are drawn from the CASP13 results pages with additional results shown for the CASP13 dataset for 'all groups' chains, scored on CASP13 PDB files, by CASP domain definitions. Contact prediction accuracies were recomputed from the group 032 and 498 submissions (as RR files), compared with the distogram predictions used by AlphaFold for CASP13 submissions. Contact prediction probabilities were obtained from the distograms by summing the probability mass in each distribution below 8 Å.

For each training sequence, we searched for and aligned to the training sequence similar protein sequences in the Uniclust30[35] dataset with HHblits[36] and used the returned MSA to generate profile features with the position-specific substitution probabilities for each residue as well as covariation features—the parameters of a regularized pseudolikelihood-trained Potts model similar to CCMpred[16]. CCMPred uses the Frobenius norm of the parameters, but we feed both this norm (1 feature) and the raw parameters (484 features) into the network for each residue pair *ij*. In addition, we provide the network with features that explicitly represent gaps and deletions in the MSA. To make the network better able to make predictions for shallow MSAs, and as a form of data augmentation, we take a sample of half the sequences from the the HHblits MSA before computing the MSA-based features. Our training set contains 10 such samples for each domain. We extract additional profile features using PSI-BLAST[37].

The distance prediction neural network was trained with the following input features (with the number of features indicated in brackets).

- Number of HHblits alignments (scalar).
- Sequence-length features: 1-hot amino acid type (21 features); profiles: PSI-BLAST (21 features), HHblits profile (22 features), non-gapped profile (21 features), HHblits bias, HMM profile (30 features), Potts model bias (22 features); deletion probability (1 feature); residue index (integer index of residue number, consecutive except for multi-segment domains, encoded as 5 least-significant bits and a scalar).
- Sequence-length-squared features: Potts model parameters (484 features, fitted with 500 iterations of gradient descent using Nesterov momentum 0.99, without sequence reweighting); Frobenius norm (1 feature); gap matrix (1 feature).

The *z*-scores were taken from the results CASP13 assessors (http://predictioncenter.org/casp13/zscores_final.cgi?formula=assessors).

Distogram prediction. The inter-residue distances are predicted by a deep neural network. The architecture is a deep two-dimensional dilated convolutional residual network. Previously, a two-dimensional residual network was used that was preceded by one-dimensional embedding layers for contact prediction[21]. Our network is two-dimensional throughout and uses 220 residual blocks[29] with dilated convolutions[38]. Each residual block, illustrated in Extended Data Fig. 1b, consists of a sequence of neural network layers[39] that interleave three batchnorm layers; two 1×1 projection layers; a 3×3 dilated convolution layer and exponential linear unit (ELU)[40] nonlinearities. Successive layers cycle through dilations of 1, 2, 4, 8 pixels to allow propagation of information quickly across the cropped region. For the final layer, a position-specific bias was used, such that the biases were indexed by residue-offset (capped at 32) and bin number.

The network is trained with stochastic gradient descent using a cross-entropy loss. The target is a quantification of the distance between the C_β atoms of the residues (or C_α for glycine). We divide the range 2–22 Å into 64 equal bins. The input to the network consists of a two-dimensional array of features in which each *i*,*j* feature is the concatenation of the one-dimensional features for both *i* and *j* as well as the two-dimensional features for *i*,*j*.

Individual training runs were cross-validated with early stopping using 27 CASP11 FM domains as a validation set. Models were selected by cross-validation on 27 CASP12 FM domains.

Neural network hyperparameters

- 7 groups of 4 blocks with 256 channels, cycling through dilations 1, 2, 4, 8.
- 48 groups of 4 blocks with 128 channels, cycling through dilations 1, 2, 4, 8.
- Optimization: synchronized stochastic gradient descent
- Batch size: batch of 4 crops on each of 8 GPU workers.
- 0.85 dropout keep probability.
- Nonlinearity: ELU.
- Learning rate: 0.06.
- Auxiliary loss weights: secondary structure: 0.005; accessible surface area: 0.001. These auxiliary losses were cut by a factor 10 after 100 000 steps.
- Learning rate decayed by 50% at 150,000, 200,000, 250,000 and 350,000 steps.
- Training time: about 5 days for 600,000 steps.

Cropped distograms. To constrain memory usage and avoid overfitting, the network was always trained and tested on 64×64 regions of the distance matrix, that is, the pairwise distances between 64 consecutive residues and another group of 64 consecutive residues. For each training domain, the entire distance matrix was split into non-overlapping 64×64 crops. By training off-diagonal crops, the interaction between residues that are further apart than 64 residues could be modelled. Each crop consisted of the distance matrix that represented the juxtaposition of two 64-residue fragments. It has previously been shown[22] that contact prediction needs only a limited context window. We note that the distance predictions close to the diagonal $i=j$, encode predictions of the local structure of the protein, and for any cropped region the distances are governed by the local structure of the two fragments represented by the *i* and *j* ranges of the crop. Augmenting the inputs with the on-diagonal two-dimensional input features that correspond to both the *i* and *j* ranges provides additional information to predict the structure of each fragment and thus the distances between them. It can be seen that if the fragment structures can be well predicted (for instance, if they are confidently predicted as helices or sheets), then the prediction of a single contact

between the fragments will strongly constrain the distances between all other pairs.

Randomizing the offset of the crops each time a domain is used in training leads to a form of data augmentation in which a single protein can generate many thousands of different training examples. This is further enhanced by adding noise proportional to the ground-truth resolution to the atom coordinates, leading to variation in the target distances. Data augmentation (MSA subsampling and coordinate noise), together with dropout[41], prevents the network from overfitting to the training data.

To predict the distance distribution for all $L \times L$ residue pairs, many 64×64 crops are combined. To avoid edge effects, several such tilings are produced with different offsets and averaged together, with a heavier weighting for the predictions near the centre of the crop. To improve accuracy further, predictions from an ensemble of four separate models, trained independently with slightly different hyperparameters, are averaged together. Extended Data Figure 2b, c shows examples of the true distances and the mode of the distogram predictions for a three-domain CASP13 target, T0990.

As the network has a rich representation capable of incorporating both profile and covariation features of the MSA, we argue that the network can be used to predict the secondary structure directly. By mean- and max- pooling the two-dimensional activations of the penultimate layer of the network separately in both i and j, we add an additional one-dimensional output head to the network that predicts eight-class secondary structure labels as computed by DSSP[42] for each residue in j and i. The resulting accuracy of the Q3 (distinguishing the three helix/sheet/coil classes) predictions is 84%, which is comparable to the state-of-the-art predictions[43]. The relative accessible surface area (ASA) of each residue can also be predicted.

The one-dimensional pooled activations are also used to predict the marginal Ramachandran distributions, $P(\varphi_i, \psi_i|S,\mathrm{MSA}(S))$, independently for each residue, as a discrete probability distribution approximated to 10° (1,296 bins). In practice during CASP13 we used distograms from a network that was trained to predict distograms, secondary structure and ASA. Torsion predictions were taken from a second similar network trained to predict distograms, secondary structure, ASA and torsions, as the former had been more thoroughly validated.

Extended Data Figure 3b shows that an important factor in the accuracy of the distograms (as has previously been found with contact prediction systems) is N_{eff}, the effective number of sequences in the MSA[20]. This is the number of sequences found in the MSA, discounting redundancy at the 62% sequence identity level, which we then divide by the number of residues in the target, and is an indication of the amount of covariation information in the MSA.

Distance potential. The distogram probabilities are estimated for discrete distance bins; therefore, to construct a differentiable potential, the distribution is interpolated with a cubic spline. Because the final bin accumulates probability mass from all distances beyond 22 Å, and as greater distances are harder to predict accurately, the potential was only fitted up to 18 Å (determined by cross-validation), with a constant extrapolation thereafter. Extended Data Figure 3c (bottom) shows the effect of varying the resolution of the distance histograms on structure accuracy.

To predict a reference distribution, a similar model is trained on the same dataset. The reference distribution is not conditioned on the sequence, but to account for the atoms between which we are predicting distances, we do provide a binary feature $\delta_{\alpha\beta}$ to indicate whether the residue is a glycine (C_α atom) or not (C_β) and the overall length of the protein.

A distance potential is created from the negative log likelihood of the distances, summed over all pairs of residues i, j (Supplementary equation (1)). With a reference state, this becomes the log-likelihood ratio of the distances under the full conditional model and under the background model (Supplementary equation (2)).

Torsions are modelled as a negative log likelihood under the predicted torsion distributions. As we have marginal distribution predictions, each of which can be multimodal, it can be difficult to jointly optimize the torsions. To unify all of the probability mass, at the cost of modelling fidelity of multimodal distributions, we fitted a unimodal von Mises distribution to the marginal predictions. This potential was summed over all residues i (Supplementary equation (3)).

Finally, to prevent steric clashes, a van der Waals term was introduced through the use of Rosetta's $V_{\mathrm{score2\_smooth}}$. Extended Data Figure 3c (top) shows the effect on the accuracy of the structure prediction of different terms in the potential.

Structure realization by gradient descent. To realize structures that minimize the constructed potential, we created a differentiable model of ideal protein backbone geometry, giving backbone atom coordinates as a function of the torsion angles (φ, ψ): $\mathbf{x} = G(\varphi, \psi)$. The complete potential to be minimized is then the sum of the distance, torsion and score2_smooth (Supplementary equation (4)). Although there is no guarantee that these potentials have equivalent scale, scaling parameters on the terms were introduced and chosen by cross-validation on CASP12 FM domains. In practice, equal weighting for all terms was found to lead to the best results.

As every term in V_{total} is differentiable with respect to the torsion angles, given an initial set of torsions φ, ψ, which can be sampled from the predicted torsion marginals, we can minimize V_{total} using a gradient descent algorithm, such as L-BFGS[31]. The optimized structure is dependent on the initial conditions, so we repeat the optimization multiple times with different initializations. A pool of the 20 lowest-potential structures is maintained and once full, we initialize 90% of trajectories from those with 30° noise added to the backbone torsions (the remaining 10% still being sampled from the predicted torsion distributions). In CASP13, we obtained 5,000 optimization runs for each chain. Figure 2c shows the change in TM score against the number of restarts per protein. As longer chains take longer to optimize, this work load was balanced across $(50 + L)/2$ parallel workers. Extended Data Figure 4 shows similar curves against computation time, always comparing sampling starting torsions from the predicted marginal distributions with restarting from the pool of previous structures.

Accuracy. We compare the final structures to the experimentally determined structures to measure their accuracy using metrics such as TM score, GDT_TS (global distance test, total score[44]) and r.m.s.d. All of these accuracy measures require geometric alignment between the candidate structure and the experimental structure. An alternative accuracy measure that requires no alignment is the lDDT[45], which measures the percentage of native pairwise distances D_{ij} under 15 Å, with sequence offsets $\geq r$ residues, that are realized in a candidate structure (as d_{ij}) within a tolerance of the true value, averaging across tolerances of 0.5, 1, 2 and 4 Å (without stereochemical checks), as shown in Supplementary equation (5)).

As the distogram predicts pairwise distances, we can introduce distogram lDDT (DLDDT), a measure similar to lDDT that is computed directly from the probabilities of the distograms, as shown in Supplementary equation (6)). As distances between residues nearby in the sequence are often short, easier to predict and are not critical in determining the overall fold topology, we set $r = 12$, considering only those distances for residues with a sequence separation ≥12. Because we predict C_β distances, for this study we computed both lDDT and DLDDT using the C_β distances. Extended Data Figure 3a shows that DLDDT_{12} has high correlation (Pearson's $r = 0.92$ for CASP13) with the lDDT_{12} of the realized structures.

Full chains without domain segmentation. Parameterizing proteins of length L by two torsion angles per residue, the dimension of the space of structures grows as $2L$; thus, searching for structures of large proteins becomes much more difficult. Traditionally this problem was addressed by splitting longer protein chains into pieces—termed domains—that fold independently. However, domain segmentation from the sequence alone is itself difficult and error-prone. For this study, we avoided domain segmentation and folded entire chains. Typically, MSAs are based on a given domain segmentation; however, we used a sliding window approach, computing a full-chain MSA to predict a baseline full-sequence distogram. We then computed MSAs for subsequences of the chain, trying windows of size 64, 128, 256 with offsets at multiples of 64. Each of these MSAs gave rise to an individual distogram that corresponded to an on-diagonal square of the full-chain distogram. We averaged all of these distograms together, weighted by the number of sequences in the MSA to produce an average full-chain distogram that is more accurate in regions in which many alignments can be found. For the CASP13 assessment, full chains were relaxed with Rosetta relax with a potential of $V_{\text{Talaris2014}} + 0.2\, V_{\text{distance}}$ (weighting determined by cross-validation) and submissions from all of the systems were ranked based on this potential.

CASP13 results. For CASP13, the five AlphaFold submissions were from three different systems, all of which used potentials based on the neural network distance predictions. The systems that are not described here are described in a separate paper[8]. Before T0975, two systems based on simulated annealing and fragment assembly (and using 40-bin distance distributions) were used. From T0975 onward, newly trained 64-bin distogram predictions were used and structures were generated by the gradient descent system described here (three independent runs) as well as one of the fragment assembly systems (five independent runs). The five submissions were chosen from these eight structures (the lowest potential structure generated by each independent run) with the first submission (top-one) being the lowest-potential structure generated by gradient descent. The remaining four submissions were the four best other structures, with the fifth being a gradient descent structure if none had been chosen for position 2, 3 or 4. All submissions for T0999 were generated by gradient descent. Extended Data Figure 5a shows the methods used for each submission, comparing with 'back-fill' structures generated by a single run of gradient descent for targets before T0975. Extended Data Figure 5b shows that the gradient descent method that was used later in CASP performed better than the fragment assembly method, in each category. Extended Data Figure 5c compares the accuracy of the AlphaFold submissions for FM and FM/TBM domains with the next best group 322. The assessors of CASP13 FM used expert visual inspection[46] to choose the best submissions for each target and found that AlphaFold had nearly twice as many best models as the next best group.

Biological relevance of AlphaFold predictions. There is a wide range of uses of predicted structures, all with different accuracy requirements, from generally understanding the fold shape to understanding detailed side-chain configurations in binding regions. Contact predictions alone can guide biological insights[47], for instance, to target mutations to destabilize the protein. Figure 1c and Extended Data Fig. 2a show that the accuracy of the contact predictions from AlphaFold exceeds that of the state-of-the-art predictions. In Extended Data Figs. 6–8, we present further results that show that the accuracy improvements of AlphaFold lead to more accurate interpretations of function (Extended Data Fig. 6); better interface prediction for protein–protein interactions (Extended Data Fig. 7); better binding pocket prediction (Extended Data Fig. 8) and improved molecular replacement in crystallography.

Thus far only template-based predictions have been able to deliver the most accurate predictions. Although AlphaFold is able to match TBM without using templates, and in some cases outperform other methods (for example, T0981-D5, 72.8 GDT_TS, and T0957s1-D2, 88.0 GDT_TS, two TBM-hard domains for which the top-one model of AlphaFold is 12 GDT_TS better than any other top-one submission), the accuracy for FM targets still lags behind that for TBM targets and can still not be relied on for the detailed understanding of hard structures. In an analysis of the performance of CASP13 TBM predictions for molecular replacement, another study[48] reported that the AlphaFold predictions (raw coordinates, without B-factors) led to a marginally greater log-likelihood gain than those of any other group, indicating that these improved structures can assist in phasing for X-ray crystallography.

Interpretation of distogram neural network. We have shown that the deep distance prediction neural network achieves high accuracy, but we would like to understand how the network arrives at its distance predictions and—in particular—to understand how the inputs to the model affect the final prediction. This might improve our understanding of the folding mechanisms or suggest improvements to the model. However, deep neural networks are complex nonlinear functions of their inputs, and so this attribution problem is difficult, under-specified and an on-going topic of research. Even so, there are a number of methods for such analysis: here we apply Integrated Gradients[49] to our trained distogram network to indicate the location of input features that affect the network's predictions of a particular distance.

In Extended Data Fig. 9, plots of summed absolute Integrated Gradient, $\sum_c |S^{I,J}_{i,j,c}|$, (defined in Supplementary equations (7)–(9)) are shown for selected I,J output pairs in T0986s2; and in Extended Data Fig. 10, the top-10 highest attribution input pairs for each output pair are shown on top of the top-one predicted structure of AlphaFold. The attribution maps are sparse and highly structured, closely reflecting the predicted geometry of the protein. For the four in-contact pairs presented (1, 2, 3, 5), all of the highest attribution pairs are pairs within or between the secondary structure that one or both of the output pair(s) are members of. In 1, the helix residues are important as well as connections between the strands that follow either end of the helix, which might indicate strain on the helix. In 2, all of the most important residue pairs connect the same two strands, whereas in 3, a mixture of inter-strand pairs and strand residues is most salient. In 5, the most important pairs involve the packing of nearby secondary structure elements to the strand and helix. For the non-contacting pair, 4, the most important input pairs are the residues that are geometrically between I and J in the predicted protein structure. Furthermore, most of the high-attribution input pairs are themselves in contact.

As the network is tasked with predicting the spatial geometry, with no structure available at the input, these patterns of interaction indicate that the network is using intermediate predictions to discover important interactions and channelling information from related residues to refine the final prediction.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Our training, validation and test data splits (CATH domain codes) are available from https://github.com/deepmind/deepmind-research/tree/master/alphafold_casp13. The following versions of public datasets were used in this study: PDB 2018-03-15; CATH 2018-03-16; Uniclust30 2017-10; and PSI-BLAST nr dataset (as of 15 December 2017).

Code availability

Source code for the distogram, reference distogram and torsion prediction neural networks, together with the neural network weights and input data for the CASP13 targets are available for research and non-commercial use at https://github.com/deepmind/deepmind-research/tree/master/alphafold_casp13. We make use of several open-source libraries to conduct our experiments, particularly HHblits[36], PSI-BLAST[37] and the machine-learning framework TensorFlow (https://github.com/tensorflow/tensorflow) along with the TensorFlow library Sonnet (https://github.com/deepmind/sonnet), which provides implementations of individual model components[50]. We also used Rosetta[9] under license.

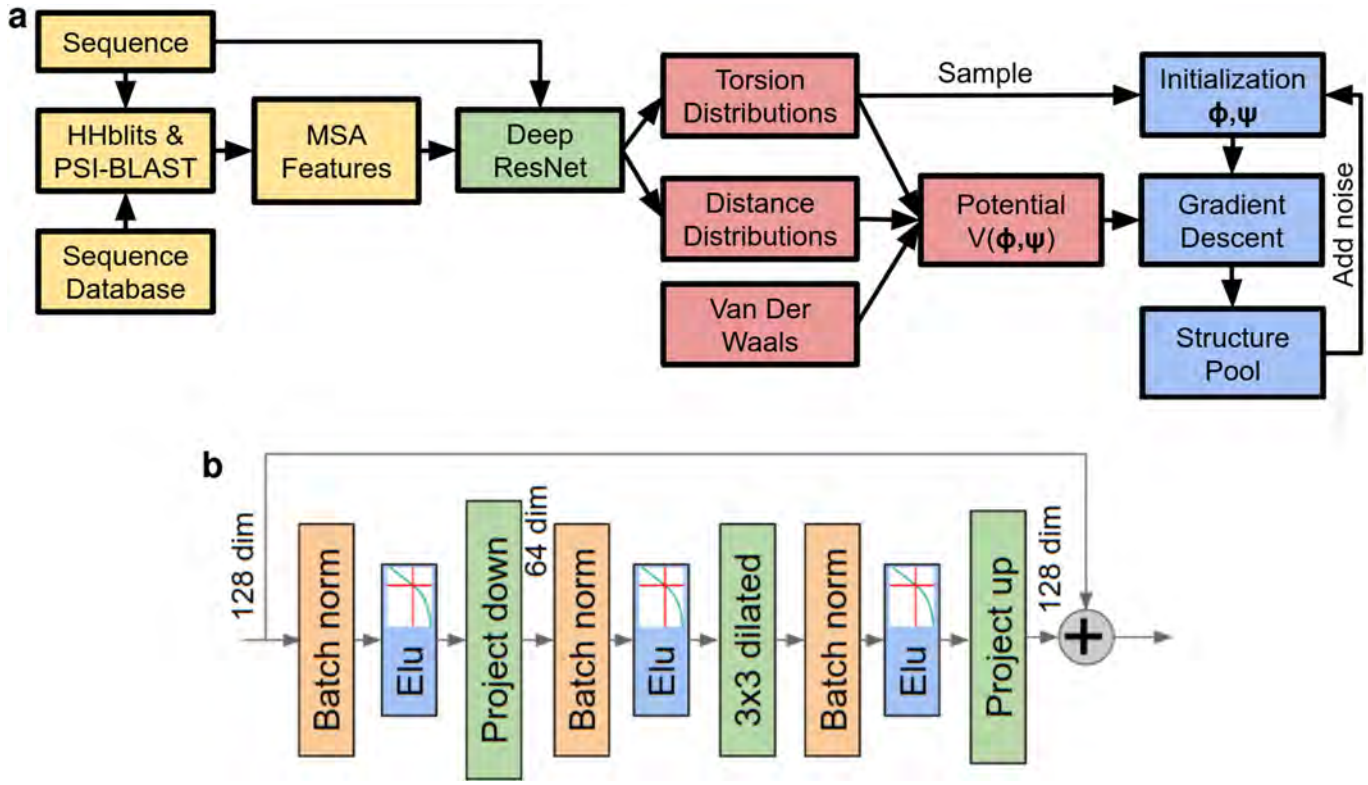
34. Dawson, N. L. et al. CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* **45**, D289–D295 (2017).
35. Mirdita, M. et al. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* **45**, D170–D176 (2017).
36. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM–HMM alignment. *Nat. Methods* **9**, 173–175 (2012).
37. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
38. Yu, F. & Koltun, V. Multi-scale context aggregation by dilated convolutions. Preprint at *arXiv* https://arxiv.org/abs/1511.07122 (2015).
39. Oord, A. d. et al. Wavenet: a generative model for raw audio. Preprint at *arXiv* https://arxiv.org/abs/1609.03499 (2016).
40. Clevert, D.-A., Unterthiner, T. & Hochreiter, S. Fast and accurate deep network learning by exponential linear units (ELUs). Preprint at *arXiv* https://arxiv.org/abs/1511.07289 (2015).
41. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
42. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
43. Yang, Y. et al. Sixty-five years of the long march in protein secondary structure prediction: the final stretch? *Briefings Bioinf.* **19**, 482–494 (2018).
44. Zemla, A., Venclovas, C., Moult, J. & Fidelis, K. Processing and analysis of CASP3 protein structure predictions. *Proteins* **37**, 22–29 (1999).
45. 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).
46. Abriata, L. A., Tamo, G. E. & Dal Peraro, M. A further leap of improvement in tertiary structure prediction in CASP13 prompts new routes for future assessments. *Proteins* **87**, 1100–1112 (2019).
47. Kayikci, M. et al. Visualization and analysis of non-covalent contacts using the Protein Contacts Atlas. *Nat. Struct. Mol. Biol.* **25**, 185–194 (2018).
48. Croll, T. I. et al. Evaluation of template-based modeling in CASP13. *Proteins* **87**, 1113–1127 (2019).
49. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *Proc. 34th International Conference on Machine Learning* **Vol. 70**, 3319–3328 (2017).
50. Abadi, M. et al. Tensorflow: a system for large-scale machine learning. In *Proc. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* 265–283 (2016).
51. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
52. Cong, Q. et al. An automatic method for CASP9 free modeling structure prediction assessment. *Bioinformatics* **27**, 3371–3378 (2011).
53. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
54. Tovchigrechko, A., Wells, C. A. & Vakser, I. A. Docking of protein models. *Protein Sci.* **11**, 1888–1896 (2002).
55. Audet, M. et al. Crystal structure of misoprostol bound to the labor inducer prostaglandin E_2 receptor. *Nat. Chem. Biol.* **15**, 11–17 (2019).

Acknowledgements We thank C. Meyer for assistance in preparing the paper; B. Coppin, O. Vinyals, M. Barwinski, R. Sun, C. Elkin, P. Dolan, M. Lai and Y. Li for their contributions and support; O. Ronneberger for reading the paper; the rest of the DeepMind team for their support; the CASP13 organisers and the experimentalists whose structures enabled the assessment.

Author contributions R.E., J.J., J.K., L.S., A.W.S., C.Q., T.G., A.Ž., A.B., H.P. and K.S. designed and built the AlphaFold system with advice from D.S., K.K. and D.H. D.T.J. provided advice and guidance on protein structure prediction methodology. S.P. contributed to software engineering. S.C., A.W.R.N., K.K. and D.H. managed the project. J.K., A.W.S., T.G., A.Ž., A.B., R.E., P.K. and J.J. analysed the CASP results for the paper. A.W.S. and J.K. wrote the paper with contributions from J.J., R.E., L.S., T.G., A.B., A.Ž., D.T.J., P.K., K.K. and D.H. A.W.S. led the team.

Competing interests A.W.S., J.K., T.G., J.J., L.S., R.E., H.P., C.Q., K.S., A.Ž. and A.B. have filed provisional patent applications relating to machine learning for predicting protein structures. The remaining authors declare no competing interests.

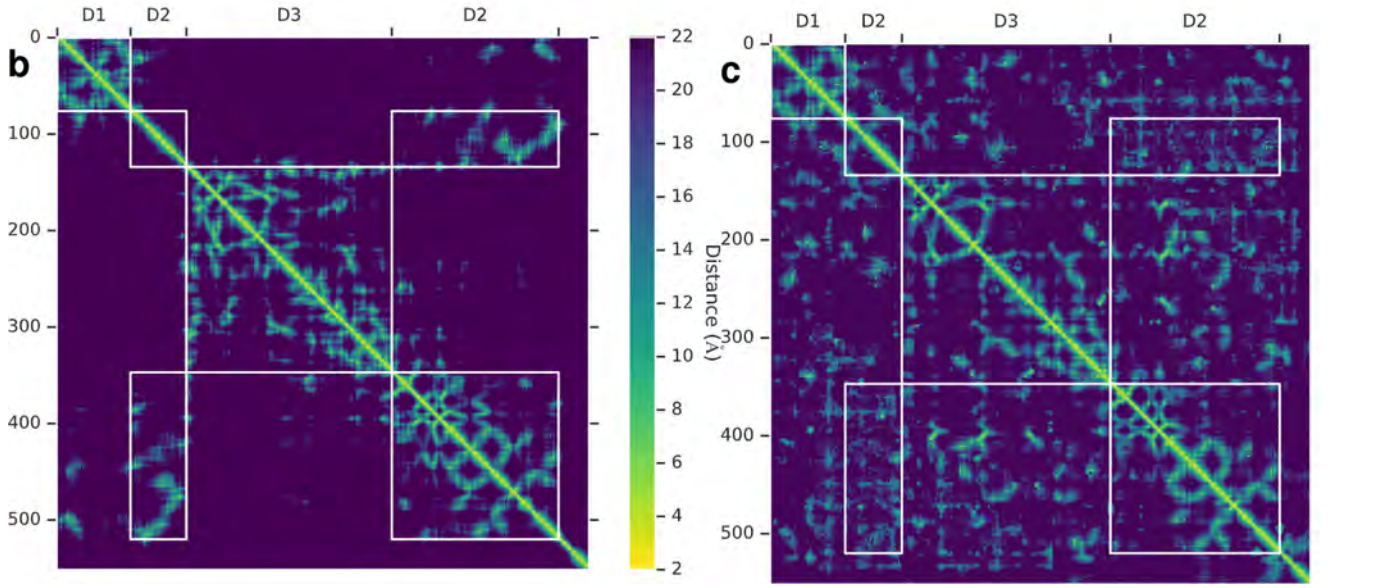
Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1923-7.
Correspondence and requests for materials should be addressed to A.W.S.
Peer review information *Nature* thanks Mohammed AlQuraishi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



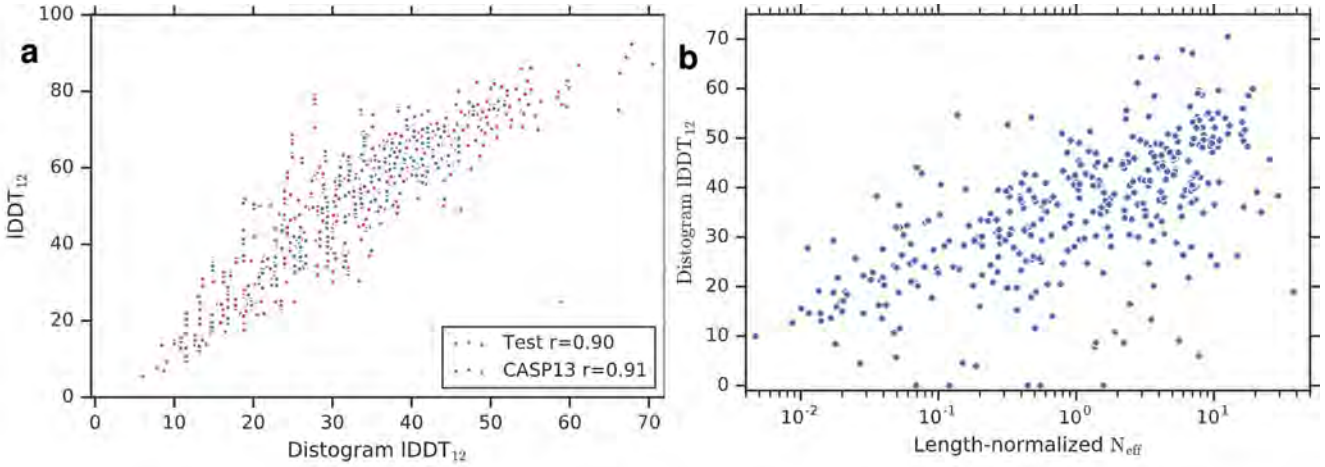
Extended Data Fig. 1 | Schematics of the folding system and neural network. **a**, The overall folding system. Feature extraction stages (constructing the MSA using sequence database search and computing MSA-based features) are shown in yellow; the structure-prediction neural network in green; potential construction in red; and structure realization in blue. **b**, The layers used in one block of the deep residual convolutional network. The dilated convolution is applied to activations of reduced dimension. The output of the block is added to the representation from the previous layer. The bypass connections of the residual network enable gradients to pass back through the network undiminished, permitting the training of very deep networks.

a

Contact precisions		L long			L/2 long			L/5 long		
Set	N	AF	498	032	AF	498	032	AF	498	032
FM	31	**46.1**	43.1	40.1	**58.5**	54.9	51.6	**69.9**	67.3	61.9
FM/TBM	12	**59.1**	53.0	48.9	**74.2**	64.5	64.2	**85.3**	81.0	79.6
TBM	61	**68.3**	65.5	61.9	**82.4**	80.3	76.4	**90.6**	90.5	87.1



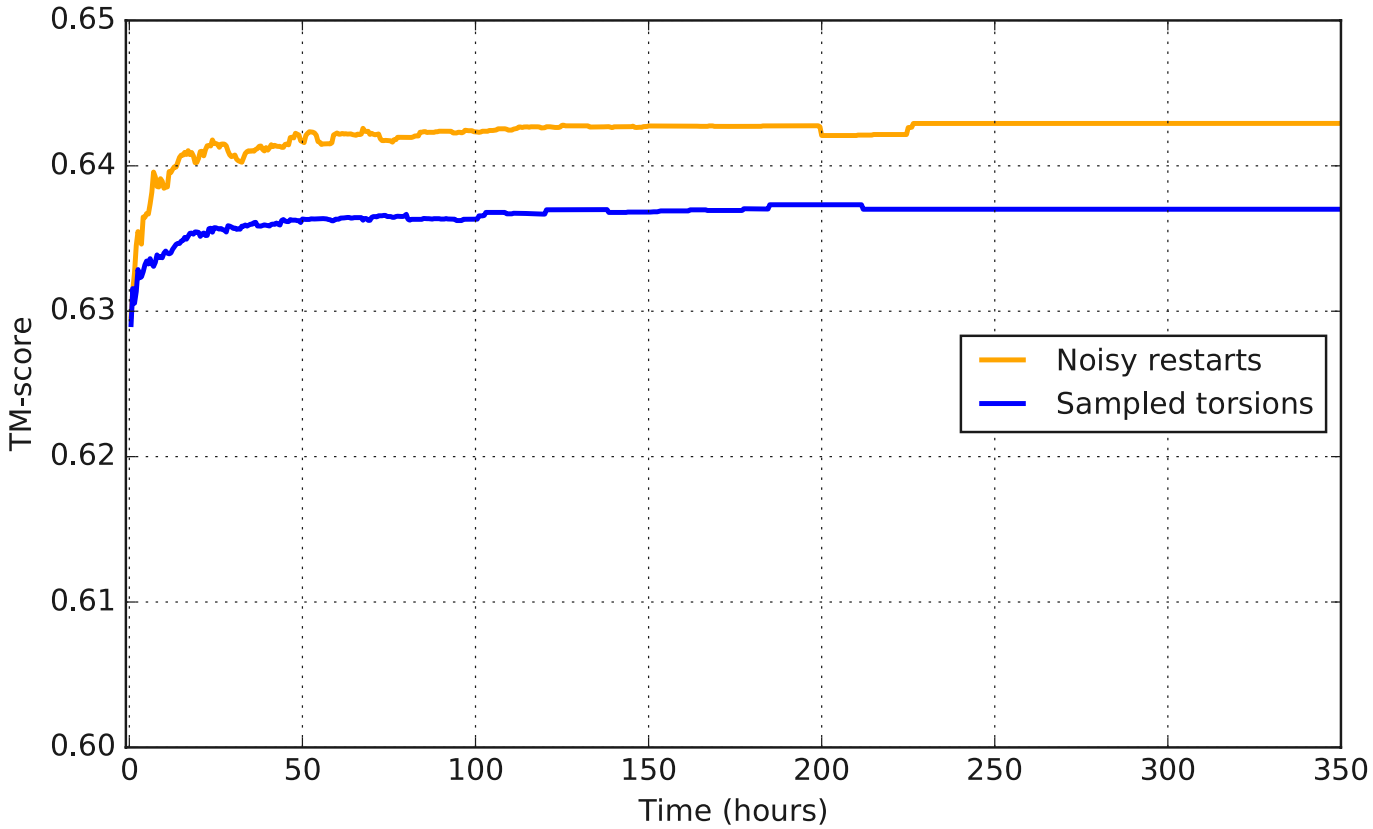
Extended Data Fig. 2 | CASP13 contact precisions. a, Precisions (as shown in Fig. 1c) for long-range contact prediction in CASP13 for the most probable L, $L/2$ or $L/5$ contacts, where L is the length of the domain. The distance distributions used by AlphaFold (AF) in CASP13, thresholded to contact predictions, are compared with submissions by the two best-ranked contact prediction methods in CASP13: 498 (RaptorX-Contact[26]) and 032 (TripletRes[32]), on 'all groups' targets, with updated domain definitions for T0953s2. **b**, **c**, True distances (**b**) and modes of the predicted distogram (**c**) for CASP13 target T0990. CASP divides this chain into three domains as shown (D3 is inserted in D2) for which there are 39, 36 and 42 HHblits alignments, respectively (from the CASP website).



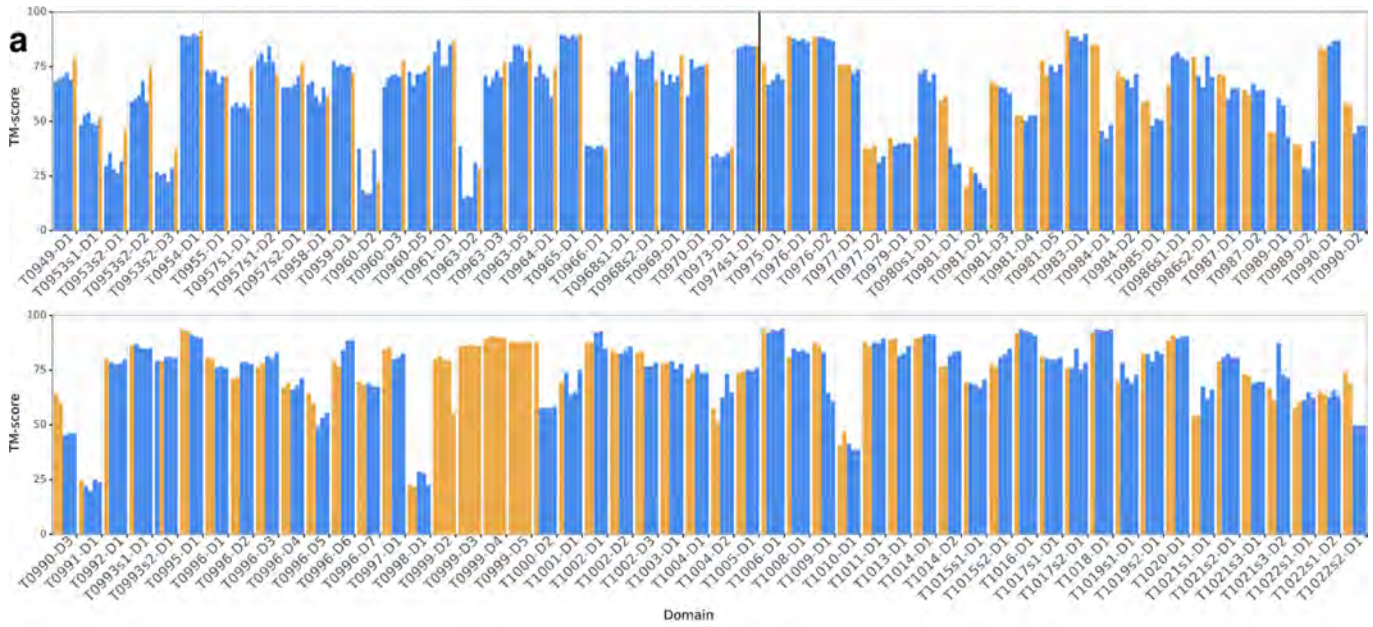
c

Potential	Bins	TM-score	GDT_TS	lDDT	RMSD (Å)	$-\log_{10} P$
Full + relax	51/64	0.649	65.8	54.2	5.94	7.3
Full	51/64	0.642	65.0	53.9	5.91	–
W/o reference	51/64	0.632	64.3	50.0	6.64	4.0
W/o `score2_smooth`	51/64	0.641	64.8	53.7	5.93	1.2
W/o torsions	51/64	0.637	64.3	53.6	6.04	8.2
W/o distogram	51/64	0.266	29.1	19.1	14.88	130
Full	48/64	0.643	65.0	54.1	5.90	
Full	24/32	0.643	65.0	53.8	5.89	
Full	12/16	0.644	65.1	53.9	5.85	
Full	6/8	0.641	64.6	53.7	5.94	
Full	3/4	0.620	62.4	52.8	6.22	
Full	2/3	0.576	58.2	49.3	8.38	

Extended Data Fig. 3 | Analysis of structure accuracies. a, $lDDT_{12}$ versus distogram $lDDT_{12}$ (see Methods, 'Accuracy'). The distogram accuracy predicts the lDDT of the realized structure well (particularly for medium- and long-range residue pairs, as well as the TM score as shown in Fig. 4a) for both CASP13 ($n = 500$: 5 decoys for domains excluding T0999) and test ($n = 377$) datasets. Data are shown with Pearson's correlation coefficients. **b**, $DLDDT_{12}$ against the effective number of sequences in the MSA (N_{eff}) normalized by sequence length ($n = 377$). The number of effective sequences correlates with this measure of distogram accuracy ($r = 0.634$). **c**, Structure accuracy measures, computed on the test set ($n = 377$), for gradient descent optimization of different forms of the potential. Top, removing terms in the potential, and showing the effect of following optimization with Rosetta relax. 'P' shows the significance of the potential giving different results from 'Full', for a two-tailed paired data *t*-test. 'Bins' shows the number of bins fitted by the spline before extrapolation and the number in the full distribution. In CASP13, splines were fitted to the first 51 of 64 bins. Bottom, reducing the resolution of the distogram distributions. The original 64-bin distogram predictions are repeatedly downsampled by a factor of 2 by summing adjacent bins, in each case with constant extrapolation beyond 18 Å (the last quarter of the bins). The two-level potential in the final row, which was designed to compare with contact predictions, is constructed by summing the probability mass below 8 Å and between 8 and 14 Å, with constant extrapolation beyond 14 Å. The TM scores in this table are plotted in Fig. 4b.

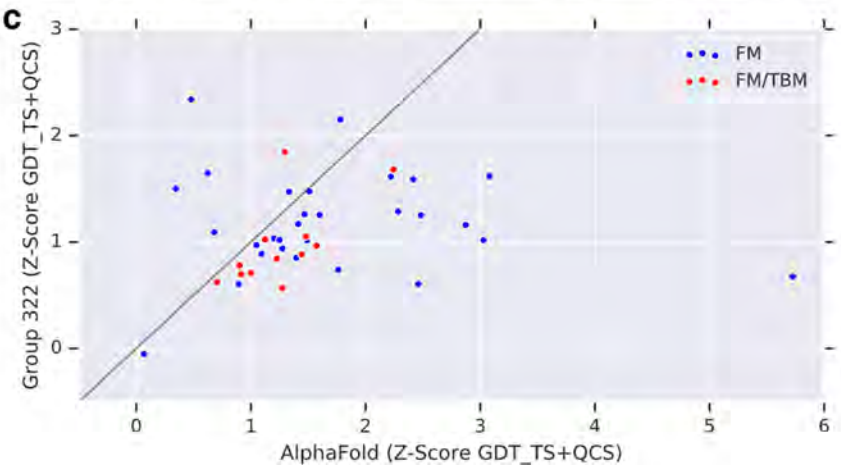


Extended Data Fig. 4 | TM score versus per-target computation time computed as an average over the test set. Structure realization requires a modest computation budget, which can be parallelized over multiple machines. Full optimization with noisy restarts (orange) is compared with initialization from sampled torsions (blue). Computation is measured as the product of the number of (CPU-based) machines and time elapsed and can be largely parallelized. Longer targets take longer to optimize. Figure 2e shows how the TM score increases with the number of repeats of gradient descent. $n = 377$.

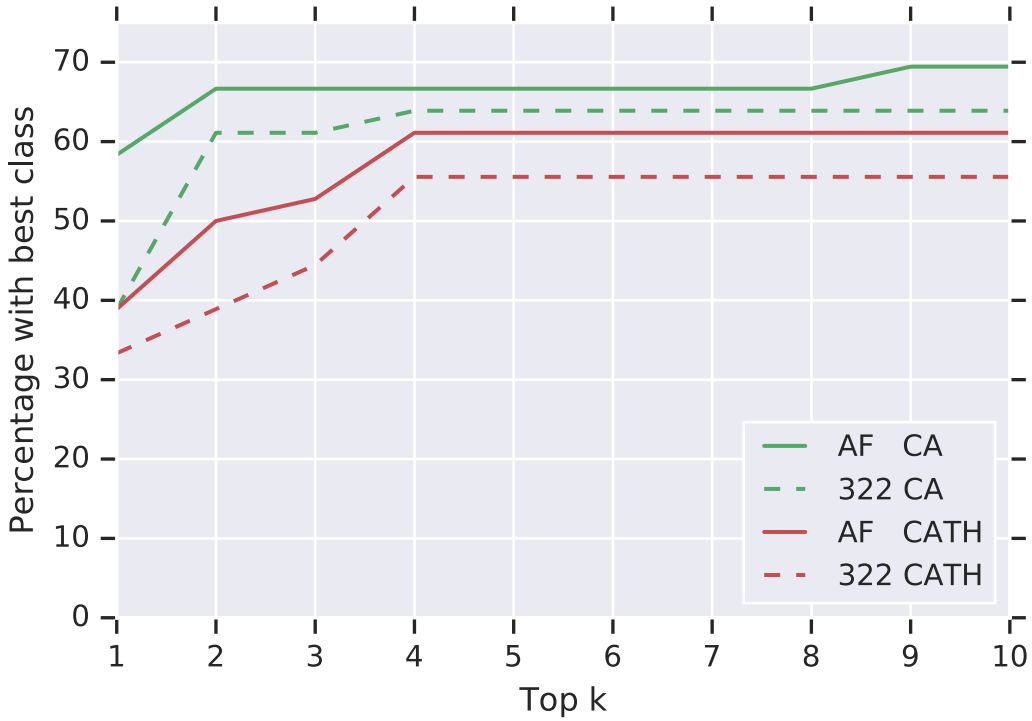


b

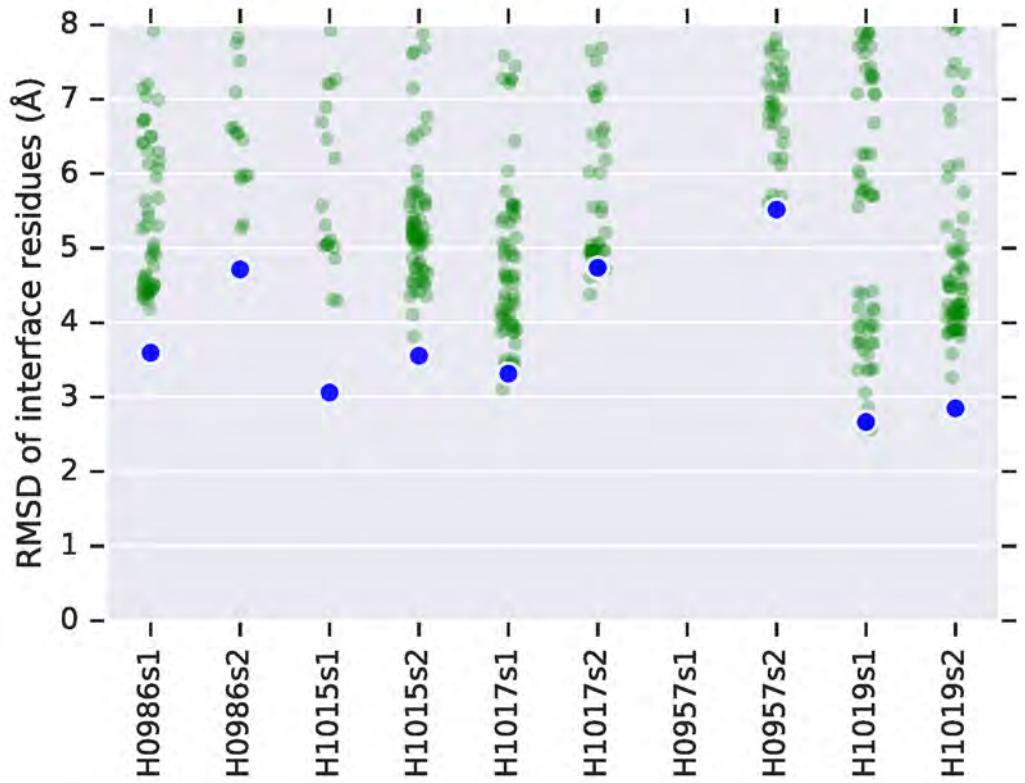
Method	FM	TBM/FM	TBM	All
Top-1	58.0	68.1	76.2	69.9
Best-of-5	62.6	73.6	78.6	73.2
1× gradient descent	58.4	71.6	76.3	70.4
1× fragment assembly	54.3	69.9	74.5	68.0



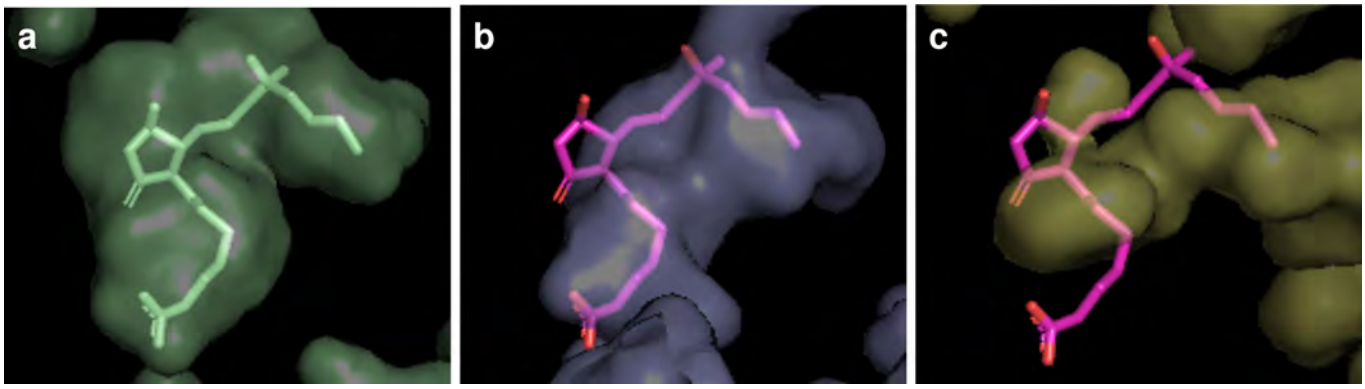
Extended Data Fig. 5 | AlphaFold CASP13 results. a, The TM score for each of the five AlphaFold CASP13 submissions are shown. Simulated annealing with fragment assembly entries are shown in blue. Gradient-descent entries are shown in yellow. Gradient descent was only used for targets T0975 and later, so to the left of the black line we also show the results for a single 'back-fill' run of gradient descent for each earlier target using the deployed system. T0999 (1,589 residues) was manually segmented based on HHpred[51] homology matching. **b**, Average TM scores of the AlphaFold CASP13 submissions (n = 104 domains), comparing the first model submitted, the best-of-five model (submission with highest GDT_TS), a single run of full-chain gradient descent (a CASP13 run for T0975 and later, back-fill for earlier targets) and a single CASP13 run of fragment assembly with domain segmentation (using a gradient descent submission for T0999). **c**, The formula-standardized (z) scores of the assessors for GDT TS + QCS[52], best-of-five for CASP FM (n = 31) and FM/TBM (n = 12) domains comparing AlphaFold with the closest competitor (group 322), coloured by domain category. AlphaFold performs better (P = 0.0032, one-tailed paired statistic t-test).



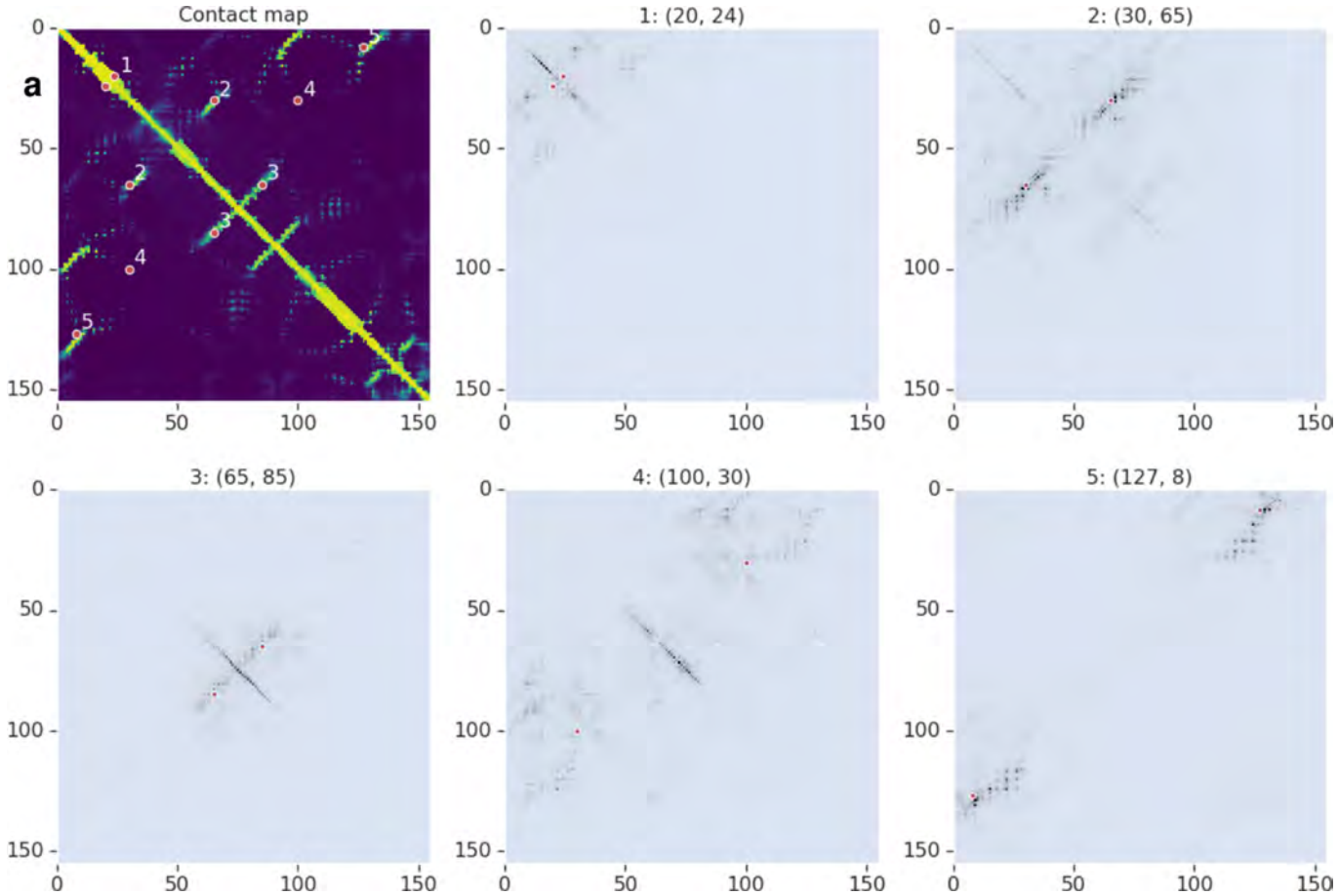
Extended Data Fig. 6 | Correct fold identification by structural search in CATH. Often protein function can be inferred by finding homologous proteins of known function. Here we show that the FM predictions of AlphaFold give greater accuracy in a structure-based search for homologous domains in the CATH database. For each of the FM or TBM/FM domains, the top-one submission and ground truth are compared to all 30,744 CATH S40 non-redundant domains with TM-align[53]. For the 36 domains for which there is a good ground-truth match (score > 0.5), we show the percentage of decoys for which a domain with the same CATH code (CATH in red, CA in green; CAT results are close to CATH results) as the top ground-truth match is in the top-*k* matches with score > 0.5. Curves are shown for AlphaFold and the next-best group (322). AlphaFold predictions determine the matching fold more accurately. Determination of the matching CATH domain can provide insights into the function of a new protein.



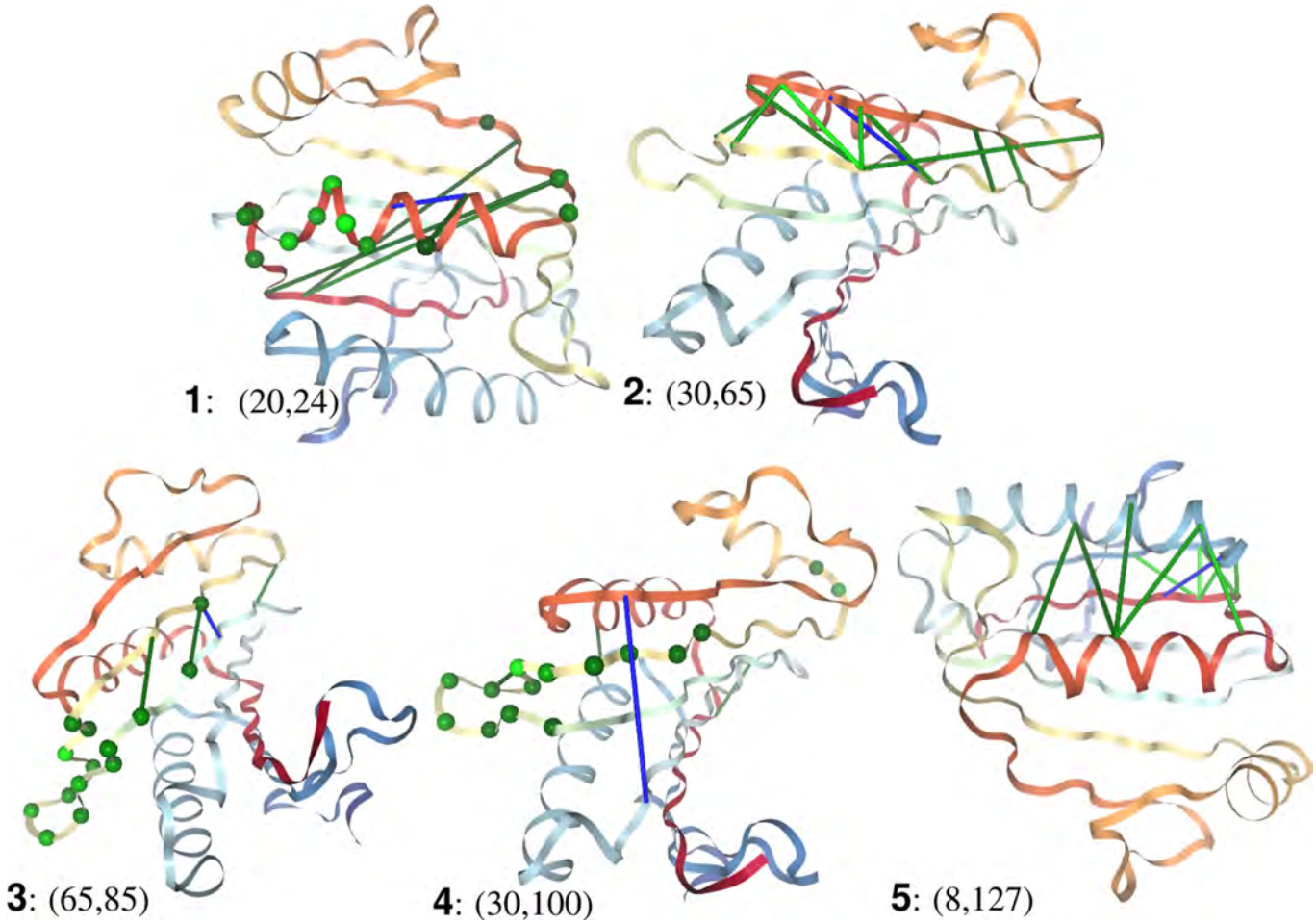
Extended Data Fig. 7 | Accuracy of predictions for interfaces. Protein–protein interaction is an important domain for understanding protein function that has hitherto largely been limited to template-based models because of the need for high-accuracy predictions, although there has been moderate success[54] in docking with predicted structures up to 6 Å r.m.s.d. This figure shows that the predictions by AlphaFold improve accuracy in the interface regions of chains in hetero-dimer structures and are probably better candidates for docking, although docking did not form part of the AlphaFold system and all submissions were for isolated chains rather than complexes. For the five all-groups heterodimer CASP13 targets, the full-atom r.m.s.d. values of the interface residues (residues with a ground-truth inter-chain heavy-atom distance <10 Å) are computed for the chain submissions of all groups (green), relative to the target complex. Results >8 Å are not shown. AlphaFold (blue) achieves consistently high accuracy interface regions and, for 4 out of 5 targets, predicts interfaces below <5 Å for both chains.



Extended Data Fig. 8 | Ligand pocket visualizations for T1011. T1011 (PDB 6M9T) is the EP3 receptor bound to misoprostol-FA[55]. **a**, The native structure showing the ligand in a pocket. **b**, **c**, Submission 5 (78.0 GDT TS) by AlphaFold (**b**), made without knowledge of the ligand, shows a pocket more similar to the true pocket than that of the best other submission (322, model 3, 68.7 GDT TS) (**c**). Both submissions are aligned to the native protein using the same subset of residues from the helices close to the ligand pocket and visualized with the interior pocket together with the native ligand position.



Extended Data Fig. 9 | Attribution map of distogram network. The contact probability map of T0986s2, and the summed absolute value of the Integrated Gradient, $\sum_c |S^{IJ}_{i,j,c}|$, of the input two-dimensional features with respect to the expected distance between five different pairs of residues (*I*,*J*): (1) a helix self-contact, (2) a long-range strand–strand contact, (3) a medium-range strand–strand contact, (4) a non-contact and (5) a very long-range strand–strand contact. Each pair is shown as two red dots on the diagrams. Darker colours indicate a higher attribution weight.



Extended Data Fig. 10 | Attribution shown on predicted structure. For T0986s2 (TM score 0.8), the top 10 input pairs, including self-pairs, with the highest attribution weight for each of the five output pairs shown in Extended Data Fig. 9, are shown as lines (or spheres for self-pairs) coloured by sensitivity, lighter green colours indicate more sensitive, and the output pair is shown as a blue line.

nature research

Corresponding author(s): Andrew Senior

Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

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

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	To ingest and pre-process the data, we used the following tools: PDB 2018-03-15 ; CATH 2018-03-16; HHblits based on version 3.0-beta.3; HHpred web server; Uniclust30 2017-10; PSI-BLAST version 2.6.0; SST web server (March 2019); BioPython v1.65; Rosetta v3.5; TM-align 20160521, as well as custom code written using Python 2.7. See the methods section for more details.
Data analysis	The networks used the TensorFlow library with custom extensions. Inference code for distance prediction networks used in CASP13 will be open-sourced. Analysis was performed with custom code written in Python 2.7. Visualizations were made with PyMol 2.2.0 software. Please see methods section for more detail.

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The following public datasets were used in this work:
- PDB 2018-03-15
- CATH 2018-03-16
- Uniclust30 2017-10
- PSI-BLAST nr dataset (as of 2017-12-15)

We will make available our train/test split (CATH domain codes).

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	Tests were carried out on two sets of domains generated by predetermined methods. The principal set consisted of 104 CASP13 domains as segmented by the CASP13 assessors. (In some figures 100 domains are used, excluding the domains of T0999 where slightly different methods were used; or 41 considering FM + TBM/FM targets) For the CASP13 datasets, accuracy measures are reported using structures & distance distributions computed during CASP (except for the "back-fill" gradient descent structures where specified). With n=41 domains, we show (Extended Data Figure 5) that (with p=0.0032) AlphaFold's results were better than the next best group. The other test set of 377 domains was extracted from PDB, chosen to be from separate homologous superfamilies (CATH code) none of which were represented in the training set. For each superfamily the exemplar was chosen at random from the s35 cluster representatives. Extended Data Table 3c shows that this sample size is sufficient to show that certain terms in the potential (but not score2) make a difference in the final GDT_TS score.
Data exclusions	The training set used one example per s35 cluster, using the representative chosen by CATH. We further balance (and reduce the size of) the test set by keeping one example per superfamily. CASP13 domains are only those "all groups" targets which were scored as part of CASP13. Some CASP13 targets were excluded by the assessors e.g. because of publications or failure to solve a structure.
Replication	Actual competition entries were used and were not replicated, but we show comparisons for top-5 decoys. Multiple networks were trained independently and were found to give consistent results (on CASP11/12 validation sets). 4 such networks, yielding similar results, were used in CASP and are made available with the source code. Structure generation is stochastic, but we show (Extended Data Fig 4) that it converges with relatively few attempts, so results are reproducible.
Randomization	Superfamilies were randomly assigned to PDB train or PDB test set. For each superfamily in the test set, a random s35 cluster representative was picked for the 377 example test set. CASP13 assessors used a manual process (blind to the investigators) to determine evaluation units and to determine whether each domain should be treated as FM, TBM or TBM/FM.
Blinding	Investigators were blind to the CASP13 targets which were sequestered during the assessment. Investigators were not blind to the PDB train/test split, but cross-validation used CASP11 & 12 domains and the main results are on the blind CASP13. Models were not retrained after defining the 377 domain set.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

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