Title

Accelerating crystal structure determination with iterative AlphaFold prediction

Journal

Acta Crystallographica Section D, Structural Biology, 79(3)

ISSN

2059-7983

Authors

Terwilliger, Thomas C

Afonine, Pavel V

Liebschner, Dorothee

et al.

Publication Date

2023-03-01

DOI

10.1107/s205979832300102x

STRUCTURAL BIOLOGY

ISSN 2059-7983

Accelerating crystal structure determination with iterative *AlphaFold* prediction

Thomas C. Terwilliger,[a,b]* Pavel V. Afonine,[c] Dorothee Liebschner,[c] Tristan I. Croll,[d] Airlie J. McCoy,[d] Robert D. Oeffner,[d] Christopher J. Williams,[e] Billy K. Poon,[c] Jane S. Richardson,[e] Randy J. Read[d] and Paul D. Adams[c,f]

[a]New Mexico Consortium, Los Alamos, NM 87544, USA, [b]Los Alamos National Laboratory, Los Alamos, NM 87545, USA, [c]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [d]Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Hills Road, Cambridge CB2 0XY, United Kingdom, [e]Department of Biochemistry, Duke University, Durham, NC 27710, USA, and [f]Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA. *Correspondence e-mail: tterwilliger@newmexicoconsortium.org

Received 30 November 2022
Accepted 3 February 2023

Edited by A. Gonzalez, Lund University, Sweden

Keywords: *AlphaFold*; artificial intelligence; automated structure determination; model building.

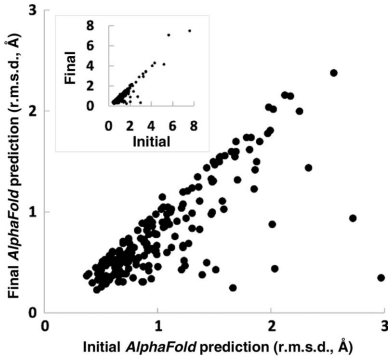
Supporting information: this article has supporting information at journals.iucr.org/d

Experimental structure determination can be accelerated with artificial intelligence (AI)-based structure-prediction methods such as *AlphaFold*. Here, an automatic procedure requiring only sequence information and crystallographic data is presented that uses *AlphaFold* predictions to produce an electron-density map and a structural model. Iterating through cycles of structure prediction is a key element of this procedure: a predicted model rebuilt in one cycle is used as a template for prediction in the next cycle. This procedure was applied to X-ray data for 215 structures released by the Protein Data Bank in a recent six-month period. In 87% of cases our procedure yielded a model with at least 50% of C^{α} atoms matching those in the deposited models within 2 Å. Predictions from the iterative template-guided prediction procedure were more accurate than those obtained without templates. It is concluded that *AlphaFold* predictions obtained based on sequence information alone are usually accurate enough to solve the crystallographic phase problem with molecular replacement, and a general strategy for macromolecular structure determination that includes AI-based prediction both as a starting point and as a method of model optimization is suggested.

1. Introduction

The development of artificial intelligence (AI)-based methods for the prediction of protein structures has been widely recognized as a turning point in structural biology (Baek *et al.*, 2021; Jumper *et al.*, 2021; Callaway, 2022; Thornton *et al.*, 2021). Predictions using *AlphaFold* (Jumper *et al.*, 2021), *RoseTTAFold* (Baek *et al.*, 2021) and related methods (Lin *et al.*, 2022) are far more accurate than previous generations of predictions (Kryshtafovych *et al.*, 2021), enabling large-scale analyses of protein function without requiring experimental structural information for each protein (van Breugel *et al.*, 2022; Thornton *et al.*, 2021). Nevertheless, there has been considerable discussion of the limitations of AI-based models (Moore *et al.*, 2022; Shao *et al.*, 2022).

The potential for using *AlphaFold* predictions to facilitate structure determination by X-ray crystallography and cryo-EM has been rapidly appreciated in the structural biology community (Akdel *et al.*, 2022; Barbarin-Bocahu & Graille, 2022; Bond & Cowtan, 2022; Chen *et al.*, 2022; Gong *et al.*, 2023; McCoy *et al.*, 2022; Medina *et al.*, 2022; Moi *et al.*, 2022; Stsiapanava *et al.*, 2022). *AlphaFold* predictions made in the CASP14 blind test of structure prediction were shown to be effective as starting models for X-ray structure determination



using the molecular-replacement method (McCoy *et al.*, 2022). *AlphaFold* models and the accompanying predicted aligned error (PAE) matrices can be used to identify domain boundaries in proteins (Oeffner *et al.*, 2022) that could be helpful in the design of trimmed versions of proteins suitable for crystallization (Lorimer *et al.*, 2009; Perrakis & Sixma, 2021).

AlphaFold predictions can include information from a template, and we recently showed that a prediction based on a template can be more accurate than either a sequence-based prediction or the template itself (Terwilliger *et al.*, 2022). This property allows iterative improvement of modeling and *AlphaFold* prediction, and here we describe an automated procedure that can accomplish this using X-ray crystallographic data.

2. Methods

2.1. Data and models from the PDB

We chose models from the PDB with the goal of obtaining a representative set of challenging structures determined after the training of *AlphaFold* (which used structures through April 2018). We selected all 215 unique protein-containing structures determined by the single-wavelength anomalous diffraction (SAD) method over a six-month period (release dates from 8 December 2021 to 29 June 2022). Crystallographic data for each model were obtained from the PDB. The X-ray data typically included the anomalous data used to solve the structures, but we did not use the anomalous information (Bijvoet pairs of reflections were averaged). Some structures contained nonprotein contents, such as waters, ions, small molecules, RNA, DNA and covalent modifications. All of these were removed from the deposited models and were not considered in our analyses. For structural comparisons, protein chains from deposited models were superimposed on our automatically generated models using crystallographic symmetry operations (with origin shifts, as appropriate).

2.2. Automatic structure determination with iterative *AlphaFold* prediction

Our procedure for automated structure determination consists of cycles of *AlphaFold* prediction, trimming and splitting predictions into compact domains, molecular replacement (in the first cycle), morphing full-length predictions onto the model obtained from molecular replacement or from a previous cycle, refinement, model rebuilding and trimming. These steps are described in detail below. At least three cycles are carried out, and the process is terminated when the rebuilt models for subsequent cycles have an overall r.m.s.d. to the previous model of less than 0.25 times the high-resolution limit of the X-ray data (controlled by the parameter `cycle_rmsd_to_resolution_ratio`). The rationale for scaling this to the resolution is that lower resolution structures are anticipated to have larger coordinate errors. The entire process is completely automated and can be carried out with the *Phenix PredictAndBuild* tool. We used default parameters in all of the structure redeterminations described here.

2.3. *AlphaFold* prediction

We used a local installation of *AlphaFold*2 (Baek *et al.*, 2021), configured as a server using software from *ColabFold* (Mirdita *et al.*, 2022), to carry out *AlphaFold* predictions using the simplified methods available in *Phenix* (Liebschner *et al.*, 2019). The sequences used in prediction were obtained from the sequence file supplied by the PDB for the corresponding structure. An alternative to using a local installation is available in *Phenix*; this alternate method uses Google *Colab* with a *Phenix* script to carry out predictions.

Predictions without templates were carried out using random seeds to initiate five *AlphaFold* predictions; the prediction with the highest average pLDDT (confidence) value was kept. Predictions with templates were carried out in the same way but supplying a template. Templates consisted of the rebuilt model obtained in a previous cycle of prediction and rebuilding. Two forms of templates were used: one containing just main-chain and C^{β} atoms and one including all side-chain atoms. For shorter chains both forms of the template were used (ten total predictions) and for longer chains only the main-chain and C^{β} atoms were included. This limitation was due to our server and the uploading method for template files: side-chain-containing templates packaged for uploading could not have more than 65 536 characters. This limitation is not present when using the *Phenix Colab* script to carry out the calculation.

2.4. *AlphaFold* model preparation

We used the *Phenix* tool *process_predicted_model* (Oeffner *et al.*, 2022) to trim residues that had lower than moderate confidence (pLDDT < 70), to convert pLDDT values to estimated atomic displacement parameters and to automatically split predicted models into domains.

2.5. Molecular replacement

We used *Phaser* (McCoy *et al.*, 2007) to carry out default molecular-replacement (MR) analyses with X-ray data and the processed *AlphaFold* predictions. The number of copies of each prediction used in MR was the number of copies of the corresponding sequence in the deposited sequence file. During MR, several values of the high-resolution limit were automatically tried and the one that yielded either a result reported as convincing by the *Phaser* software (McCoy *et al.*, 2007) or the highest value of the log-likelihood gain was used. The high-resolution limit of the data used after molecular replacement was the resolution obtained from the PDB entry.

After the first cycle, the trimmed rebuilt model from the previous cycle was used in place of a model from MR.

2.6. Morphing predicted models onto rebuilt models

Full-length predicted models were morphed (distorted) to match the model obtained from MR using the *Phenix superpose_and_morph* tool (Terwilliger *et al.*, 2022) and the `direct_morph` keyword. This tool automatically identifies parts of the predicted model that match the target model,

superimposes these parts and then smoothly deforms the model between the superposed parts. This morphing procedure creates plausible models when the required distortion is small (in the range of a few Å). However, when the distortion is large the models can be highly implausible.

2.7. Refinement

The *phenix.refine* tool (Afonine *et al.*, 2012) was used for refinement with default values for all parameters, except that the checks for overlapping atoms and long bond lengths were disabled to allow refinement to proceed even if an implausible model was encountered.

2.8. Model rebuilding

The models obtained after MR and refinement were automatically rebuilt with the *Phenix AutoBuild* tool (Terwilliger *et al.*, 2008), which carries out iterative crystallographic rebuilding, refinement and density modification to yield a rebuilt model and a density-modified map. The resulting density-modified map was then used in a second stage of model rebuilding in which the model obtained from MR and refinement was rebuilt using real-space *Phenix* tools developed for models from cryo-electron microscopy experiments (Terwilliger *et al.*, 2022). After carrying out these two approaches, the model that had the lowest free *R* value was used in subsequent steps.

2.9. Model trimming to match maps

We trimmed the rebuilt models and docked *AlphaFold* predictions to provide hypotheses for further structure-determination steps that only include the parts of the model that are likely to be correct. The basis for choosing which segments (sequential sets of residues) in a model to keep included map–model comparisons and, for predicted models, the predicted model confidence. Map–model comparisons consisted of comparing each residue in a model with the corresponding density map (typically the density-modified map from model rebuilding) and calculating the local map–model correlation (the correlation between map values and those in a map calculated from the model). Predicted model confidence consisted of the pLDDT values from *AlphaFold* prediction.

The first trimming step consisted of removing segments in which the local map correlation and (for predicted models) the pLDDT values are below their corresponding cutoff values. The map correlation and the pLDDT values were smoothed over a window of typically ten residues (controlled by the parameter `minimum_domain_length`). The cutoff values are calculated from the mean and standard deviation of the highest half of the correlation (or pLDDT) values, where the cutoff is typically the mean minus three times the standard deviation (controlled by the parameter `cc_sd_ratio`).

The second trimming step consisted of removing residues at the segment ends that have a map correlation and a smoothed map correlation below a higher cutoff (typically the mean minus twice the standard deviation, controlled by the parameter `cc_sd_ratio_end`). Segments that are shorter than the length of the smoothing window are then removed.

Finally, segments with a much lower mean map correlation than most segments are removed. This is achieved by calculating the average map correlation for each segment. The mean and standard deviation of the top half of these average map correlations are then calculated. A cutoff is then calculated as the higher of the mean correlation scaled by a constant with a typical value of 0.64 (given by the square of the parameter `reasonable_cc_ratio`) and the mean correlation minus a constant with a typical value of 0.3 (given by twice the parameter `reasonable_cc_di`). All segments with a mean map correlation below this cutoff are removed and a new model is created from all remaining segments.

2.10. Construction of docked *AlphaFold* predictions

In order to create a docked model in which each chain was identical to an *AlphaFold* prediction, the chains in a rebuilt model were used to guide the positioning of *AlphaFold* predictions. These docked predicted models are not always geometrically plausible, as the *AlphaFold* predictions are not necessarily the same as the corresponding structures. Rather, these docked predicted models are a convenient vehicle for managing the *AlphaFold* predictions for a structure, and in some cases they are a reasonable representation of the structure.

2.11. Model comparisons and map–model correlations

We compared models that were not previously superposed using least-squares superposition and calculation of r.m.s. differences in C^{α} positions with the *Phenix* (Liebschner *et al.*, 2019) tool *superpose_pdbs*. Models that were already placed in appropriate crystallographic positions were compared using automatic mapping of positions based on space-group symmetry using the *Phenix resolve* tool with the `compare_pdb` keyword.

Map–model correlations were calculated with the *Phenix* tool *get_cc_mtz_pdb*, which maximizes local map–model correlation by (i) adjusting the radius used for masking the map around each atom and (ii) modifying side-chain atomic displacement factors by adding an incremental value for each atom beyond the C^{α} atom. The correlations reported are global (using the entire map).

2.12. Map and model display

Figures were prepared with *ChimeraX* (version 1.2.5; Pettersen *et al.*, 2021).

2.13. Data and code availability

Input data for deposited models were taken from the Protein Data Bank. All models are downloadable from the PDB with links such as https://files.rcsb.org/download/7tzp.pdb or (for larger models that are not available in this format) https://files.rcsb.org/download/7tzp.cif. We used the *Phenix* tool *fetch_pdb* to download models and crystallographic data for each structure.

Predicted models, rebuilt models and density-modified map coefficients are available at https://phenix-online.org/phenix_data, along with a spreadsheet that contains all of the raw data and analyses described here. The directory `terwilliger/alphafold_crystallography_2022/` contains a README file describing the contents of the site, the spreadsheet and a `data/` directory with one compressed archive for each structure containing models and crystallographic data files. This directory also contains a compressed archive (`alphafold_crystallography.tgz`) containing all of the data and all of the scripts used to create the spreadsheet.

All code for the *Phenix* version of the *AlphaFold*2 *Colab* is freely available on GitHub at https://github.com/phenix-project/Colabs. All code for *Phenix* is available at https://phenix-online.org.

3. Results

3.1. Crystallographic structure determination using iterative *AlphaFold* prediction and rebuilding

Our iterative procedure for macromolecular structure determination by X-ray crystallography uses *AlphaFold* predictions in an initial structure-solution cycle, followed by cycles of *AlphaFold* prediction and model rebuilding in which rebuilt models from one cycle are used as templates for prediction in the next (Terwilliger, 2003; McCoy *et al.*, 2007; Terwilliger *et al.*, 2022; see Section 2). This procedure requires processed crystallographic data (structure-factor amplitudes and their uncertainties, space group and unit-cell dimensions for the crystal, and the resolution of the data) and information about the sequences of the macromolecules that are present in the crystal. It produces an optimized (density-modified; Terwilliger, 2000) electron-density map and a model based on that map. During the procedure the map itself improves as the model improves, in contrast to the related workflow for single-particle cryo-electron microscopy (Terwilliger *et al.*, 2022), where the map does not change. The approach can be used if the majority or all of a structure consists of protein. If other components are present, such as RNA/DNA, only the protein part of the resulting map is interpreted.

3.2. Determination of challenging structures using *AlphaFold* prediction

The method of molecular replacement (Rossmann, 1990), in which an initial model that is similar to the structure to be determined is used as a hypothesis for the actual structure, has been applied in about 80% of recent macromolecular crystal structure determinations (Wang *et al.*, 2017). Once the orientation and location of the initial model have been found, the model and density map are usually improved by cycles of map calculation alternating with map representation as an atomic model with restrained geometry (Perrakis *et al.*, 1999). These procedures work best if the initial models are within about 1–2 Å (r.m.s.d. of C^{α} atoms) over about 50% of the structure (Abergel, 2013). Excitement about *AlphaFold* predictions in the crystallographic field comes from the observation that such predictions may generally be accurate enough to provide the necessary starting point for molecular replacement, largely removing the need to use anomalous scattering or other experiments to obtain crystallographic phases (McCoy *et al.*, 2022; Millán *et al.*, 2021; Akdel *et al.*, 2022).

Here, we use an automated procedure to test the iterative use of *AlphaFold* predictions in macromolecular crystallography. To recreate a situation where challenging new structures are being determined using *AlphaFold* predictions, we selected structures obtained with anomalous scattering, an approach that is typically used when molecular replacement is expected to fail, but we did not include the anomalous diffraction information. For entries in the Protein Data Bank (wwPDB Consortium, 2018) released in the six-month period from 8 December 2021 to 29 June 2022 this selection yielded 215 unique structures with resolutions ranging from 1.0 to 4.6 Å.

We applied our iterative procedure for *AlphaFold* prediction and model improvement to each of the 215 deposited data sets. In seven cases the initial molecular-replacement step did not yield any solution and the analysis was not continued. For the remaining 208 data sets our procedure generated density-modified (Terwilliger, 2000) electron-density maps and models interpreting these maps. To identify which maps and models were at least partially correct, we compared the density-modified electron-density maps with model-based ($F_{\rm calc}$) maps calculated from the corresponding deposited structures. Using a conservative minimum map correlation of 0.5 as a threshold (Oeffner *et al.*, 2013), 187 of 215 analyses (87%) were successful and the remaining 28 (including the seven that failed in molecular replacement) were unsuccessful.

Fig. 1(*a*) shows the distribution of map–model correlation values. For the successful cases, the average map correlation was 0.84. The solid bars in Fig. 1(*b*) illustrate the completeness of these structures, as measured by the percentage of C^{α} atoms in deposited models matching those in the rebuilt models within 2 Å (Terwilliger *et al.*, 2022). The average completeness was 90% and all but one of the models were at least 50% complete. For the unsuccessful analyses (open bars), the completeness of the models was much lower (average of 20%). In a few of the successful instances the final structures contained domains that matched the deposited model but the connectivity between the domains was incorrect (for example PDB entry 7e1d, which is a domain-swapped dimer; Bennett *et al.*, 1994). We included all matching parts by using space-group symmetry-related copies of each chain from the deposited models in the comparisons. The overall high success rate shows that in most cases *AlphaFold* predictions are accurate and complete enough to be used as starting models in macromolecular crystallography.

AlphaFold provides residue-level confidence estimates including a predicted local difference distance test (pLDDT) associated with each amino-acid residue (Jumper *et al.*, 2021). Here, we refer to residues with pLDDT values of 90 or higher as high confidence and those with values of 70 or higher as moderate-to-high confidence (Jumper *et al.*, 2021). The success

of automated structure determination using *AlphaFold* prediction was strongly dependent on the percentage of moderate-to-high confidence residues. For successful cases, an average of 90% of residues in the *AlphaFold* predictions were predicted with moderate-to-high confidence, and the smallest percentage of moderate-to-high confidence residues was 30%. For the 28 unsuccessful cases, the average percentage predicted with moderate-to-high confidence was much lower (50%).

We assessed the similarity of predicted models and the corresponding deposited structures in more detail by superimposing the models, removing residues with low confidence (pLDDT < 70) and further removing residues where the r.m.s.d. (smoothed with a window of ten residues) was greater than 3 Å. We then noted the r.m.s.d. between matching C^{α} atoms and the coverage, defined here as the percentage of residues in the deposited model that were matched by the predicted model. For successful cases, the average C^{α} r.m.s.d. was 1.2 Å and the average coverage was 89%. Such values are normally associated with success in molecular replacement (Abergel, 2013). For unsuccessful cases the average r.m.s.d. was 2.8 Å, which was much too high to expect success in molecular replacement, and the average coverage was 50%, a borderline value for success. Overall, 19 of the 28 failures had a coverage less than 50%, an r.m.s.d. over 2 Å, or both.

When determining a new structure, metrics that do not depend on knowing the true structure are important for evaluating whether structure determination has been successful. Fig. 1(*c*) shows two metrics that are available after structure determination and that are independent of knowledge of the true structure (McCoy *et al.*, 2007). Log-likelihood gain (LLG) scores reflect the confidence of placing the model and free *R* values reflect the cross-validated agreement of the model with crystallographic data after model rebuilding and refinement. The median LLG score for successful structure determinations with *AlphaFold* predictions was about 1000, while for unsuccessful cases the median was 80. We note that in a large-scale test of molecular replacement with homology models from the PDB (Oeffner *et al.*, 2018) only 7% of cases had an LLG score over 1000 (median LLG score of 270). Our procedure using *AlphaFold* predictions therefore led to much higher LLG scores than are typically obtained with homology models. The median free *R* value for successful cases was 0.30, a value that is normally associated with a largely correct but

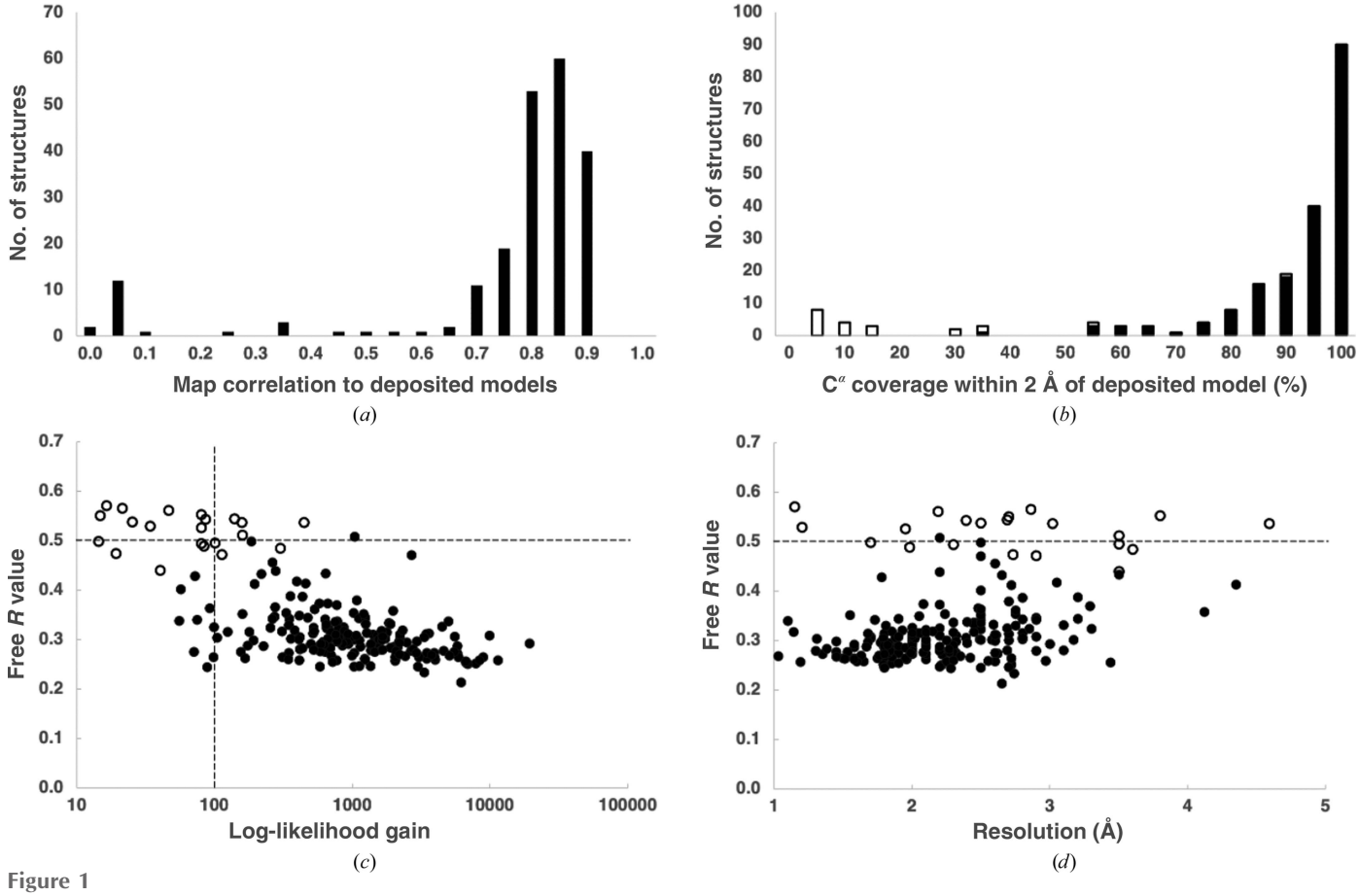


Figure 1
Structure redeterminations using *AlphaFold* predictions obtained during automated structure redeterminations for the 208 analyses that yielded a molecular-replacement solution. (*a*) Correlation of maps calculated from deposited models to density-modified maps (Terwilliger, 2000). (*b*) Percentage of C^{α} atoms in redetermined models within 2 Å of a C^{α} atom in deposited models after applying overall space-group-specific origin shifts and including space-group symmetry-related copies in the comparison. Successful cases are shown as solid bars and unsuccessful cases are shown as open bars. (*c*) Log-likelihood gain in molecular replacement and free *R* values for automatic structure analyses. Unsuccessful cases are indicated by open circles. (*d*) Free *R* values as in (*c*) but shown as a function of the high-resolution limit of the X-ray data.

unfinished structure (a structure that does not contain waters, ligands or any components other than protein). Fig. 1(*c*) illustrates that most unsuccessful solutions had free *R* values above 0.5, log-likelihood gain scores below 100, or both, suggesting that the free *R* value and log-likelihood gain can be used effectively to evaluate potential solutions in new structure determinations using this procedure, as is standard practice (McCoy *et al.*, 2007). Fig. 1(*d*) displays the free *R* values as a function of the high-resolution limit of the X-ray data and illustrates that the range of free *R* values depends slightly on the resolution of the data.

We evaluated whether our procedure may have yielded partially correct solutions for some of the cases that we considered unsuccessful. For example, the model obtained by automated structure determination for PDB entry 7f05 had a map–model correlation of just 0.24 and thus was considered not successful. Supplementary Fig. S1(*a*) shows that much of the rebuilt chain *B* for this structure (dark brown) very closely matches chain *A* in the deposited model (symmetry copy from the deposited model shown in light blue). However, the rest of the rebuilt model does not match the deposited structure. Supplementary Fig. S1(*b*) shows that the density-modified map is also quite clear. This example shows that even an unsuccessful structure can contain some correct information. Furthermore, although our procedure did not complete the structure, other approaches such as repeating molecular replacement using this one chain as a starting point (McCoy *et al.*, 2007) might well do so.

3.3. Iteration of *AlphaFold* prediction and model optimization

Our procedure for automatic structure determination iterates through *AlphaFold* prediction, using the rebuilt chains obtained at the end of one cycle of iteration as templates for *AlphaFold* prediction in the next. In this procedure, the map and model quality can improve at two stages. Firstly, the new *AlphaFold* prediction using the previously rebuilt model as a template can yield an improved predicted model (Terwilliger *et al.*, 2022), and secondly, rebuilding can improve both the model and the map (Perrakis *et al.*, 1999). Iteration is carried out until the change in the model from one cycle to the next is small. As models are not necessarily improved by rebuilding, they are evaluated based on their free *R* value, *i.e.* they are kept if the free *R* value decreases.

Fig. 2 shows this procedure applied to PDB entry 7oa7 (Shahin *et al.*, 2022), which includes X-ray data to a resolution of 1.45 Å. This figure follows the models and their fit to the map through two cycles of the procedure. The left-hand column shows ribbon overviews of superpositions onto the deposited model PDB entry 7oa7 (always in cyan). The two right-hand columns are close-ups superimposed on the map at the relevant stage: the central column (Figs. 2*b*, 2*e* and 2*h*) shows the progression for a region that started out matching the deposited structure quite well and the right-hand column (Figs. 2*c*, 2*f* and 2*i*) shows the progression for a region that started out matching quite poorly. For these two right-hand columns the predicted models are in magenta and the rebuilt models are in dark blue, and time moves vertically downward. The top row shows the initial no-template prediction (magenta) and the map output from molecular replacement. The central row shows the rebuilt (dark blue) model and the map as rebuilt in cycle 1. The bottom row shows the cycle 2 predicted model (magenta) that was given the rebuilt cycle 1 model as a template and the cycle 2 map.

An initial *AlphaFold* prediction obtained using the sequence but no structural templates is shown in Fig. 2(*a*). Some parts of the predicted model (magenta) match the deposited model (cyan) quite closely, such as the domain on the right side of Fig. 2(*a*) and the region near Ile166 and Val193. Other parts, such as the region near Tyr21 and Pro141, have different backbone conformations. Still other parts, such as much of the domain on the left side of Fig. 2(*a*), differ by a combination of local conformations and overall rotation and translation. Overall, the predicted model differs from the deposited model by a C^{α} r.m.s.d. of 2.9 Å. The initial density map obtained after molecular replacement is quite clear in the region of Ile166 and Val193 where the predicted model matches the deposited model closely (Fig. 2*b*). In contrast, the map is less clear and does not match the predicted model in the region of Tyr21 and Pro141 where the deposited and predicted models differ (Fig. 2*c*).

After automated crystallographic rebuilding starting with the predicted model shown in Fig. 2(*a*), the rebuilt model (dark blue) improves substantially (Fig. 2*d*); the overall C^{α} r.m.s.d. to the deposited model is 0.1 Å and 90% of C^{α} atoms in the deposited model are within 2 Å of those in the rebuilt model. The density map is also considerably improved (Figs. 2*e* and 2*f*) and matches the rebuilt model quite closely in the region near Ile166 and Val193 and, most notably, in the region near Tyr21 and Pro141 where the predicted and deposited models differed in the previous step.

Using the rebuilt model (Fig. 2*d*) as a template, a new *AlphaFold* prediction was obtained that is very similar to the deposited model (Fig. 2*g*; overall C^{α} r.m.s.d. of 0.5 Å) and is quite different from the initial *AlphaFold* prediction (Fig. 2*a*). Figs. 2(*h*) and 2(*i*) show that the template-based prediction matches the density-modified map both in the region near Ile166 and Val193 where the initial predicted model matched the deposited model and in the region near Tyr21 and Pro141 where it did not. The template-based prediction is even more complete than the template, with 99% of residues in the deposited model matched within 2 Å by a C^{α} atom in the prediction. (The model used as a template had only 90% matched residues.) Note that both the polypeptide backbone and the side chains in the template-based prediction closely match the density map (Fig. 2*i*) and corresponding side-chain positions in the rebuilt model (compare Figs. 2*f* and 2*i*).

Two additional cycles of rebuilding based on the model in Fig. 2(*c*) resulted in a 95% complete model, *i.e.* where 95% of the C^{α} atoms in the deposited model matched those within 2 Å in the rebuilt model, with an overall C^{α} r.m.s.d. of 0.1 Å.

In summary, although the initial *AlphaFold* prediction obtained for PDB entry 7oa7 was substantially different from

the deposited model, a structure could be obtained in the first cycle of our procedure with standard molecular-replacement and rebuilding procedures. Using the rebuilt model as a template for *AlphaFold* prediction, a new prediction was obtained that was far more accurate than the initial prediction. Three cycles of iteration for PDB entry 7oa7 led to a nearly complete model that matched the deposited structure closely and required approximately 7 h using four processors on a Linux server.

Although the template-based *AlphaFold* prediction for PDB entry 7oa7 is more complete than the rebuilt model, it is slightly less accurate (C^{α} r.m.s.d. with the deposited model of 0.5 and 0.1 Å for the *AlphaFold* and rebuilt models, respectively). This is not surprising, as the *AlphaFold* prediction uses the experimental data only indirectly (in the form of the rebuilt model) and has not been adjusted to match the density map. In a real case, both the rebuilt model and the final *AlphaFold* prediction would have utility in subsequent stages of structure determination, as the rebuilt model is slightly more accurate and the *AlphaFold* model is more complete.

Fig. 3(*a*) shows histograms of free *R* values in the first and last cycles of our procedure for all 187 successful analyses. The free *R* value decreased with iteration in 88 of 187 cases, and on average it was reduced from 0.33 to 0.31. Iteration was more successful when the free *R* value after the first cycle was poor. For the 109 cases where the free *R* value in the first cycle is

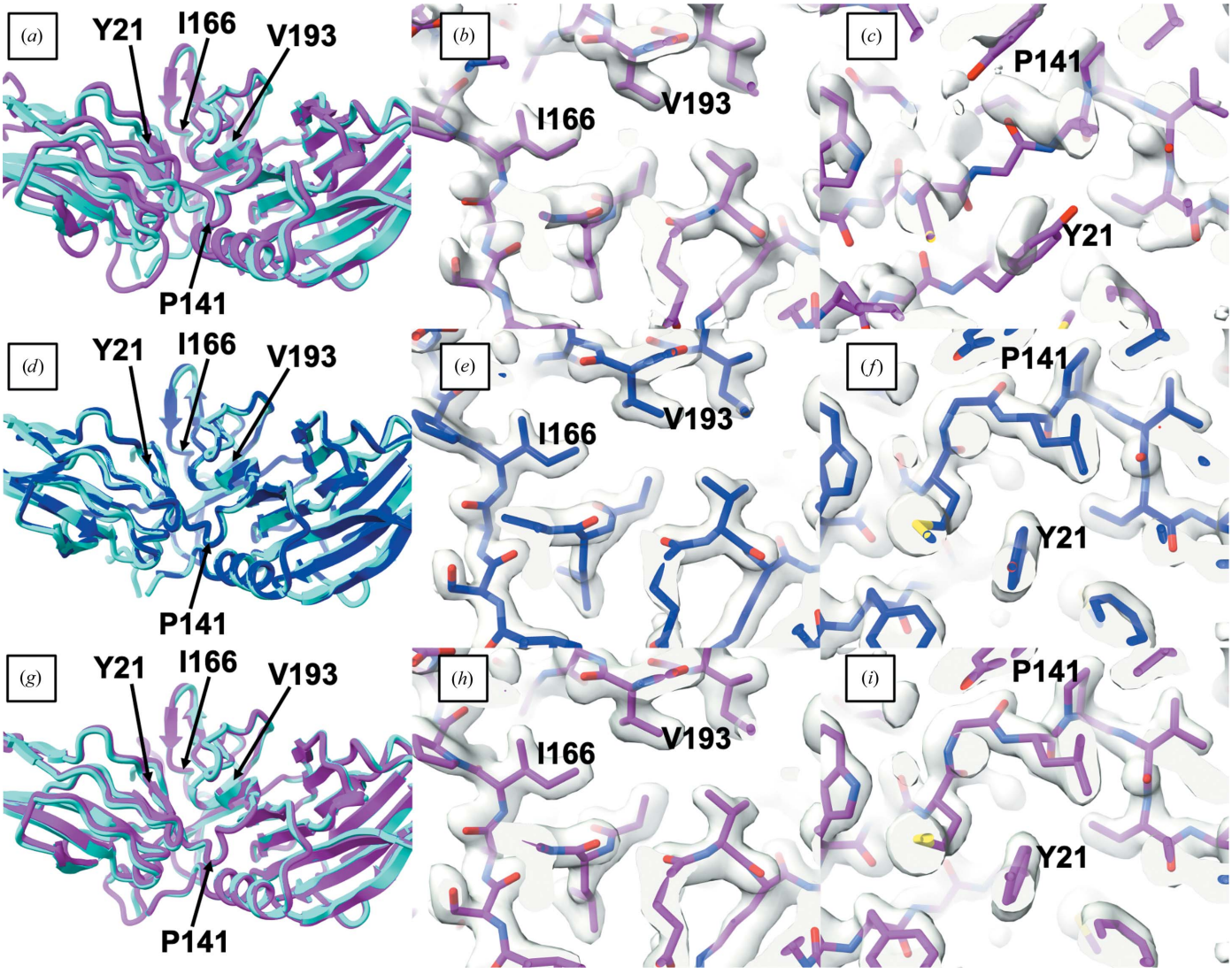


Figure 2
Iterative *AlphaFold* prediction and model rebuilding with X-ray data for PDB entry 7oa7. (*a*) Superposition of the *AlphaFold* prediction without templates (magenta) and the deposited model (cyan). The region comprising Ile166 and Val193 and the domain on the right are similar in the deposited and predicted models, while the region of Tyr21 and Pro141 and the domain shown on the left differ between the deposited and predicted models. (*b*, *c*) Close-up views of the *AlphaFold* prediction without templates superimposed on the density map obtained after molecular replacement and refinement. The good region containing Ile166 and Val193 is shown in (*b*) and the poor region containing Tyr21 and Pro141 is shown in (*c*). (*d*) Superposition of the deposited model and the model obtained by molecular replacement and automated model rebuilding using the *AlphaFold* prediction shown in (*a*) (dark blue). (*e*, *f*) Close-up of the model in (*d*) and the density-modified map obtained from the first cycle of automated model rebuilding. (*g*) Superposition of the deposited model and the superposed *AlphaFold* prediction obtained using the model in (*b*) as a template. (*h*, *i*) Close-up of the predicted model in (*g*) and the density-modified map obtained in the second iteration of model rebuilding.

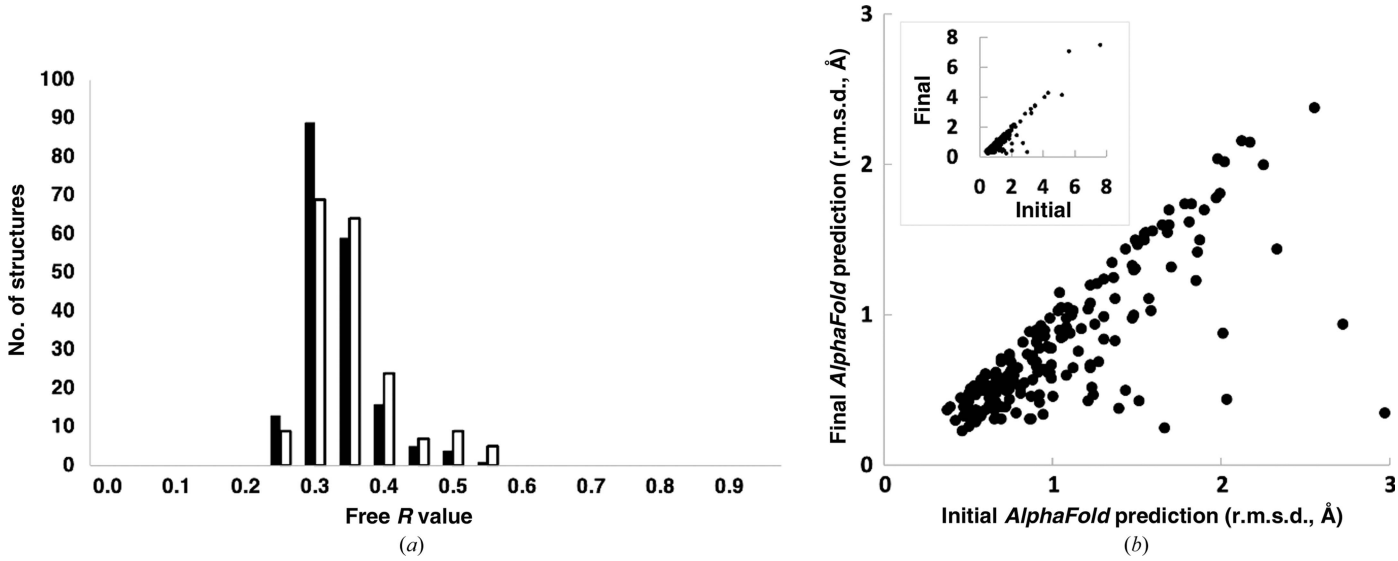


Figure 3
Iterative *AlphaFold* prediction and model rebuilding with X-ray data. (*a*) Free *R* values for the rebuilt model after the first (open bars) or last (closed bars) iteration of prediction and rebuilding (using the cycle with the lowest free *R* value as the last iteration). (*b*) R.m.s. differences between the first chain in each deposited model and the corresponding initial *AlphaFold* predictions (carried out without templates at the beginning of the first cycle; abscissa) and final *AlphaFold* predictions (carried out with templates in the last cycle; ordinate). One PDB entry (7dz9) is not shown as no rebuilt model was obtained for the first chain in the PDB entry so that the *AlphaFold* models in each cycle were the same. The inset includes nine cases that are not shown in the main panel where the initial r.m.s.d. was greater than 3 Å.

worse than 0.30, 66 were improved by iteration, while for the 78 cases where the free *R* value was better or equal to 0.30 only 22 were improved. Presumably this is because the procedures used here do not generally yield free *R* values much better than 0.30.

Fig. 3(*b*) shows the r.m.s.d. between C^{α}-atom coordinates in *AlphaFold* predictions for each structure and those in the corresponding chains of the deposited models. Most points are below the diagonal, indicating that the final *AlphaFold* prediction was more accurate than the initial prediction. For the initial *AlphaFold* predictions the median r.m.s.d. was 1.0 Å, while for the final predictions with templates it was reduced to 0.7 Å.

Overall, Fig. 3 shows that the agreement of rebuilt models with experimental data was generally very good at the end of the first cycle (mean free *R* of 0.33) and improved slightly with iteration (mean free *R* of 0.31). The accuracy of the initial *AlphaFold* predictions was also good, and improved substantially with iteration using rebuilt models as templates.

3.4. Characteristics of rebuilt and predicted models

We examined the geometric characteristics of four sets of models: initial *AlphaFold* predictions without templates, final-cycle *AlphaFold* predictions that used a template, the final rebuilt model and the deposited model. Figs. 4(*a*), 4(*b*) and 4(*c*) show comparisons between the initial *AlphaFold* prediction without templates and the final rebuilt model, and Figs. 4(*d*), 4(*e*) and 4(*f*) show comparisons between the *AlphaFold* predictions with templates and the deposited model. Fig. 4(*a*) shows the distribution of percentages of residues in favored Ramachandran conformations (Williams *et al.*, 2018). The *AlphaFold* predictions (open bars) have a mean percentage of 98%, while the rebuilt models (solid bars) average to 96%. The percentages of rotamer outliers (Williams *et al.*, 2018; Fig. 4*b*) are 0.3% and 2% for *AlphaFold* (open bars) and rebuilt models (solid bars), respectively. Clashes between nonbonded atoms were somewhat worse in the *AlphaFold* predictions [Fig. 4*c*; open bars, mean clashscore (Chen *et al.*, 2010) of 29] than for rebuilt models (solid bars, mean clashscore of 10). Fig. 4(*d*) shows the percentage of favored Ramachandran conformations for deposited models (closed bars, mean of 97%) and *AlphaFold* predictions using templates (open bars, mean of 98%). Fig. 4(*e*) shows rotamer outliers for deposited (closed bars, mean of 2%) and *AlphaFold* predictions with templates (open bars, 0.3%), and Fig. 4(*f*) shows clashscore values for deposited models (closed bars, mean of 6) and *AlphaFold* predictions with templates (open bars, mean of 27). Overall, the *AlphaFold* predictions (with or without templates) have somewhat more favorable main-chain Ramachandran conformations and the fewest rotamer outliers of the four sets of models, while the deposited models have the best clashscore values.

4. Discussion

Our analysis shows that *AlphaFold* predictions obtained based on sequence information alone are usually accurate enough to solve the crystallographic phase problem with molecular replacement. As we have shown previously, partially complete rebuilt models, such as those obtained by automatic model building, can be used effectively as templates to guide subsequent *AlphaFold* predictions, and the resulting predictions can be more accurate than either the template or predictions made without templates (Terwilliger *et al.*, 2022). Finally, we find that if a relatively accurate model is used as a template for

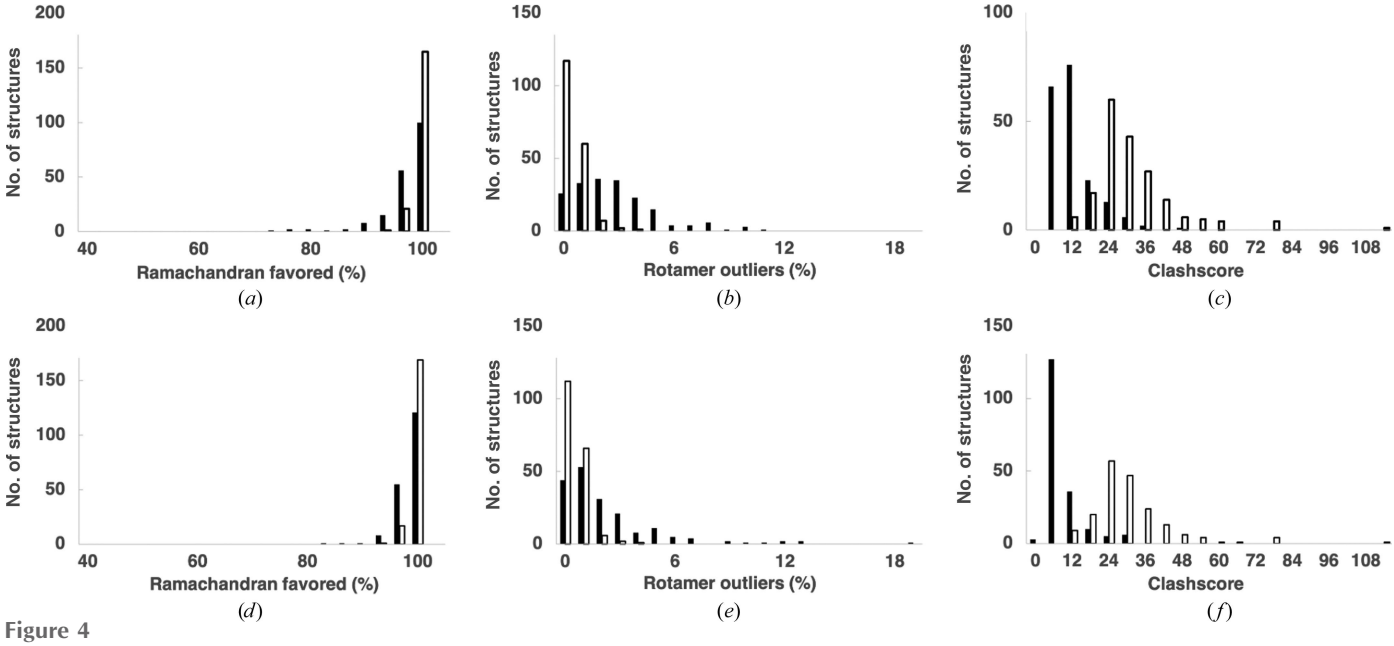


Figure 4
Geometric and packing characteristics of predicted and rebuilt models. (*a*) Histogram of the percentage of residues with 'favored' backbone conformations (open bars, *AlphaFold* predictions obtained without templates; closed bars, rebuilt models). Calculations for *AlphaFold* predictions include residues with a pLDDT of 70 or above. (*b*) Histograms of rotamer outliers (unlikely side-chain conformations). (*c*) Histograms of clashscore (Chen *et al.*, 2010) values. (*d*, *e*, *f*) As (*a*, *b*, *c*) except that the closed bars represent deposited models and the open bars represent *AlphaFold* models with templates (final cycle of iteration, trimmed to remove residues not matching map or with a pLDDT of less than 70; see Section 2).

AlphaFold prediction, the resulting predicted model can maintain both the overall conformation of the template and the details of side-chain conformations.

4.1. Strategy for crystal structure determination using template-guided *AlphaFold* prediction

These observations suggest how the overall strategy for macromolecular crystal structure determination can be revolutionized with AI-based predictions. In this new strategy, *AlphaFold* predictions are considered to be hypotheses for the actual structure. These predictions have confidence measures reflecting the expected accuracy for each residue in the prediction. Initially, the information available for structure prediction consists of the sequence of each chain and multiple sequence alignments based on these sequences. As structural information is accumulated, new predictions are made that incorporate this information.

This new strategy for macromolecular crystal structure determination starts at the planning stage and continues through the stage of obtaining a final model. In the planning stages of an experiment, confidence measures for *AlphaFold* predictions can be used to assess whether predicted models obtained using sequence alone will be useful as starting points for structure determination. Also, the *AlphaFold* predictions can be used to design trimmed versions of a macromolecule that can be successfully crystallized (Perrakis & Sixma, 2021).

To begin structure solution, *AlphaFold* predictions, trimmed to remove low-confidence regions, can be used as search models in molecular replacement, with a high probability of yielding a molecular-replacement solution that is at least partially correct.

As in a conventional structure-determination workflow, the next stage consists of refinement of the molecular-replacement solution, followed by iterations of map calculation, density modification and model rebuilding. This stage is typically carried out by existing automatic procedures that can improve both the model and the accuracy of the density maps (Perrakis *et al.*, 1999). If there are major components of the structure that are not protein (*i.e.* RNA or DNA), these components would need to be built based on the density map, yielding a more complete model and an improved density map as well. In our procedure, such additional components are automatically used in subsequent cycles.

Once an improved model has been obtained, AI-based prediction can be used again. This time, a working model is used as a template to guide *AlphaFold* prediction, as demonstrated here. This step can be thought of as a method of model optimization using *AlphaFold*. The working model guides *AlphaFold* prediction, yielding a prediction that has information from the working model but potentially with improved geometry or a more accurate overall conformation. This template-guided *AlphaFold* prediction can be carried out with or without information from multiple sequence alignments. The optimized models obtained in this way can be quite accurate, but as they are obtained without direct use of experimental information they may not exactly match the density map at this point.

The new set of *AlphaFold* predictions can then be superimposed on corresponding parts of the working model, refined using experimental data and then considered as new hypoth-

eses about the structure. It is anticipated that after a typical application of the procedure described here, both the working model and the refined new *AlphaFold* predictions would be compared with the density map, and a new composite model incorporating the best parts of each would be created. This composite model could be used as the starting point for another iteration of model rebuilding, or if it is sufficiently complete it could be used as a starting point for adding ligands, waters and other small molecules and covalent modifications. If the resolution of the data is high, this model might also be modified to explicitly add alternate conformations.

This overall strategy for structure determination differs from long-established procedures at three important steps: (i) at the planning stage, where *AlphaFold* predictions can give an idea of how challenging an experiment will be and where they can guide the design of crystallization constructs (Perrakis & Sixma, 2021); (ii) in the initial structure-determination stage, where trimmed *AlphaFold* predictions can be used in molecular replacement (McCoy *et al.*, 2022; Barbarin-Bocahu & Graille, 2022; Akdel *et al.*, 2022); and (iii) in the model-optimization stage, where template-guided *AlphaFold* prediction may be able to improve models continuously in the analysis (Terwilliger *et al.*, 2022).

A central part of this overall strategy is making use of the synergy between model rebuilding and *AlphaFold* prediction. Using a model obtained in one cycle as a template for *AlphaFold* in the next cycle often leads to improvement in the *AlphaFold* prediction, which can result in better model building and density maps in the next cycle as well. Iteration of this procedure therefore can yield improved *AlphaFold* predictions, final rebuilt models and density maps. According to our analysis, improvement of *AlphaFold* prediction is more pronounced (Fig. 3*b*) than the improvement in density maps (as reflected by the free *R* value; Fig. 3*a*). This is likely due to our use of long-established and very effective procedures for iterative model rebuilding with macromolecular crystallographic data during the first cycle, so that subsequent cycles have less effect on the density map and have a larger effect on the *AlphaFold* predictions.

4.2. Template-guided *AlphaFold* prediction in crystallography and cryo-electron microscopy

We have previously taken advantage of iterating *AlphaFold* prediction and model rebuilding in the interpretation of density maps from cryo-electron microscopy (cryo-EM; Terwilliger *et al.*, 2022). The synergy between *AlphaFold* prediction and model building is similar in the crystallographic and cryo-EM cases, but not identical. The major difference is that in the case of crystallographic data the density map can improve dramatically, while with cryo-EM data the density map is generally fixed. Although it might be anticipated that this would result in better *AlphaFold* predictions and working models by iteration with crystallographic data, our tests show that the improvement is similar for the two types of experimental methods. For crystallographic data, the median r.m.s.d. between *AlphaFold* predictions and matching PDB entries was reduced from 1.0 to 0.7 Å by iteration; for cryo-EM structures, the initial median r.m.s.d. was higher (2.5 Å) but was reduced proportionally to 1.6 Å by iteration. It seems possible that the limiting step in both cases may be model rebuilding, and with recent developments in AI-based map interpretation (Jamali *et al.*, 2022) this limitation might be greatly reduced, potentially leading to greater improvement than already obtained by iteration.

Our tests of an automatic procedure for initial crystallographic structure determination demonstrate that this overall strategy is likely to be generally effective, as most of the challenging structures that we analyzed could successfully be redetermined.

Funding information

The authors appreciate funding from Lawrence Berkeley National Laboratory (grant DE-AC02-05CH11231 to PDA), from the Phenix Industrial Consortium (to PDA), from the National Institutes of Health (grant GM063210 to PDA, RJR, TCT and JSR) and from the Wellcome Trust (grant 209407/Z/17/Z to RJR).

References

Abergel, C. (2013). *Acta Cryst.* D**69**, 2167–2173.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Akdel, M., Pires, D. E. V., Pardo, E. P., Jänes, J., Zalevsky, A. O., Mészáros, B., Bryant, P., Good, L. L., Laskowski, R. A., Pozzati, G., Shenoy, A., Zhu, W., Kundrotas, P., Serra, V. R., Rodrigues, C. H. M., Dunham, A. S., Burke, D., Borkakoti, N., Velankar, S., Frost, A., Basquin, J., Lindorff-Larsen, K., Bateman, A., Kajava, A. V., Valencia, A., Ovchinnikov, S., Durairaj, J., Ascher, D. B., Thornton, J. M., Davey, N. E., Stein, A., Elofsson, A., Croll, T. I. & Beltrao, P. (2022). *Nat. Struct. Mol. Biol.* **29**, 1056–1067.
Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J. & Baker, D. (2021). *Science*, **373**, 871–876.
Barbarin-Bocahu, I. & Graille, M. (2022). *Acta Cryst.* D**78**, 517–531.
Bennett, M. J., Choe, S. & Eisenberg, D. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 3127–3131.
Bond, P. S. & Cowtan, K. D. (2022). *Acta Cryst.* D**78**, 1090–1098.
Breugel, M. van, Rosa e Silva, I. & Andreeva, A. (2022). *Commun. Biol.* **5**, 312.
Callaway, E. (2022). *Nature*, **608**, 15–16.
Chen, L., Tan, L. & Im, Y. J. (2022). *Acta Cryst.* D**78**, 853–864.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Gong, Z., Wang, W., El Omari, K., Lebedev, A. A., Clarke, O. B. & Hendrickson, W. A. (2023). *Proc. Natl Acad. Sci. USA*, **120**, e2218630120.
Jamali, K., Kimanius, D. & Scheres, S. H. W. (2022). *arXiv*:2210.00006.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J.,

Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021). *Nature*, **596**, 583–589.
Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. (2021). *Proteins*, **89**, 1607–1617.
Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.
Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S. & Rives, A. (2022). *bioRxiv*, 2022.07.20.500902.
Lorimer, D., Raymond, A., Walchli, J., Mixon, M., Barrow, A., Wallace, E., Grice, R., Burgin, A. & Stewart, L. (2009). *BMC Biotechnol.* **9**, 36.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
McCoy, A. J., Sammito, M. D. & Read, R. J. (2022). *Acta Cryst.* D**78**, 1–13.
Medina, A., Jiménez, E., Caballero, I., Castellví, A., Triviño Valls, J., Alcorlo, M., Molina, R., Hermoso, J. A., Sammito, M. D., Borges, R. & Usón, I. (2022). *Acta Cryst.* D**78**, 1283–1293.
Millán, C., Keegan, R. M., Pereira, J., Sammito, M. D., Simpkin, A. J., McCoy, A. J., Lupas, A. N., Hartmann, M. D., Rigden, D. J. & Read, R. J. (2021). *Proteins*, **89**, 1752–1769.
Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. & Steinegger, M. (2022). *Nat. Methods*, **19**, 679–682.
Moi, D., Nishio, S., Li, X., Valansi, C., Langleib, M., Brukman, N. G., Flyak, K., Dessimoz, C., de Sanctis, D., Tunyasuvunakool, K., Jumper, J., Graña, M., Romero, H., Aguilar, P. S., Jovine, L. & Podbilewicz, B. (2022). *Nat. Commun.* **13**, 3880.
Moore, P. B., Hendrickson, W. A., Henderson, R. & Brunger, A. T. (2022). *Science*, **375**, 507.
Oeffner, R. D., Afonine, P. V., Millán, C., Sammito, M., Usón, I., Read, R. J. & McCoy, A. J. (2018). *Acta Cryst.* D**74**, 245–255.
Oeffner, R. D., Bunkóczi, G., McCoy, A. J. & Read, R. J. (2013). *Acta Cryst.* D**69**, 2209–2215.
Oeffner, R. D., Croll, T. I., Millán, C., Poon, B. K., Schlicksup, C. J., Read, R. J. & Terwilliger, T. C. (2022). *Acta Cryst.* D**78**, 1303–1314.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nat. Struct. Biol.* **6**, 458–463.
Perrakis, A. & Sixma, T. K. (2021). *EMBO Rep.* **22**, e54046.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. (2021). *Protein Sci.* **30**, 70–82.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Shahin, M., Sheppard, D., Raynaud, C., Berry, J.-L., Gurung, I., Silva, L. M., Feizi, T., Liu, Y. & Pelicic, V. (2022). *bioRxiv*, 2022.08.25.505150.
Shao, C., Bittrich, S., Wang, S. & Burley, S. K. (2022). *Structure*, **30**, 1385–1394.
Stsiapanava, A., Xu, C., Nishio, S., Han, L., Yamakawa, N., Carroni, M., Tunyasuvunakool, K., Jumper, J., de Sanctis, D., Wu, B. & Jovine, L. (2022). *Nat. Struct. Mol. Biol.* **29**, 190–193.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 1174–1182.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Terwilliger, T. C., Poon, B. K., Afonine, P. V., Schlicksup, C. J., Croll, T. I., Millán, C., Richardson, J. S., Read, R. J. & Adams, P. D. (2022). *Nat. Methods*, **19**, 1376–1382.
Thornton, J. M., Laskowski, R. A. & Borkakoti, N. (2021). *Nat. Med.* **27**, 1666–1669.
Wang, Y., Virtanen, J., Xue, Z. & Zhang, Y. (2017). *Nucleic Acids Res.* **45**, W429–W434.
Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B. III, Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (2018). *Protein Sci.* **27**, 293–315.
wwPDB Consortium (2018). *Nucleic Acids Res.* **47**, D520–D528.