Integrative Structural Biology in the Era of Accurate Structure Prediction

Gal Masrati [1], Meytal Landau [2,3], Nir Ben-Tal [1], Andrei Lupas [4]*, Mickey Kosloff [5]* and Jan Kosinski [3,6,7]*

*1 - **Department of Biochemistry and Molecular Biology,** George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 6997801, Israel*
*2 - **Department of Biology,** Technion-Israel Institute of Technology, Haifa 3200003, Israel*
*3 - **European Molecular Biology Laboratory (EMBL),** Hamburg 22607, Germany*
*4 - **Department of Protein Evolution,** Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany*
*5 - **Department of Human Biology,** Faculty of Natural Sciences, University of Haifa, 199 Aba Khoushy Ave., Mt. Carmel, 3498838 Haifa, Israel*
*6 - **Centre for Structural Systems Biology (CSSB),** Hamburg 22607, Germany*
*7 - **Structural and Computational Biology Unit,** European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany*

Correspondence to Andrei Lupas, Mickey Kosloff and Jan Kosinski: Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany (A. Lupas); University of Haifa, Haifa, 3498838, Israel (M. Kosloff); European Molecular Biology Laboratory (EMBL), Hamburg 22607, Germany (J. Kosinski). *andrei.lupas@tuebingen.mpg.de (A. Lupas), kosloff@sci.haifa.ac.il (M. Kosloff), jan.kosinski@embl.de (J. Kosinski)*
https://doi.org/10.1016/j.jmb.2021.167127

Abstract

Characterizing the three-dimensional structure of macromolecules is central to understanding their function. Traditionally, structures of proteins and their complexes have been determined using experimental techniques such as X-ray crystallography, NMR, or cryo-electron microscopy—applied individually or in an integrative manner. Meanwhile, however, computational methods for protein structure prediction have been improving their accuracy, gradually, then suddenly, with the breakthrough advance by AlphaFold2, whose models of monomeric proteins are often as accurate as experimental structures. This breakthrough foreshadows a new era of computational methods that can build accurate models for most monomeric proteins. Here, we envision how such accurate modeling methods can combine with experimental structural biology techniques, enhancing integrative structural biology. We highlight the challenges that arise when considering multiple structural conformations, protein complexes, and polymorphic assemblies. These challenges will motivate further developments, both in modeling programs and in methods to solve experimental structures, towards better and quicker investigation of structure–function relationships.

The era of accurate structure prediction

The impressive performance of AlphaFold2 in CASP14, which assesses the state-of-the-art of methods of protein structure modeling,[1] demonstrated that methods for protein structure prediction based on deep learning have reached the point where they can produce models of similar quality to experimental structures.[2–4] Although at this time (June 2021), AlphaFold2 is not yet available as a

public service, the authors declared to release the software soon.[5] In the meantime, RoseTTAFold, a new program of the Rosetta suite that builds on deep learning algorithms used in AlphaFold2, reported much-improved accuracy and has been recently released within the Robetta server.[6] We surmise that implementation of these algorithms into other leading structure prediction servers, such as Quark,[7] I-TASSER,[8] and tFold (https://drug.ai.tencent.com/console/en/tfold), is ongoing. Indeed, these and other servers already showed substantial improvements in accuracy between CASP13 and CASP14 by using deep learning, suggesting that publicly available methods will rival the performance of AlphaFold2. It is thus reasonable to assume that AlphaFold2 and publicly available servers will provide timely and free, or nearly free, accurate modeling services to the academic community. In this light, we foresee increased interplay between this new generation of modeling methods and established experimental methods. We posit that accurate protein structure prediction will open new possibilities in integrative structural biology. We discuss how experimental methodologies might be affected by new modeling methods (Figure 1), how modeling could be combined with experimental methods, and what current gaps and challenges warrant changes in these tools to better enable future progress.

Modeling and experiment – historical perspective

Proteins are essential agents of all living systems and for their activity they mostly need to assume defined three-dimensional structures.[9] Efforts to determine these structures at atomic resolution date back to the 1920s, when the newly discovered X-rays were used for fiber diffraction studies on a broad range of biological materials, including wool, horn, nails, muscle, and tendons.[10] These data led to the high-resolution description of protein secondary structure in 1950[11,12] and soon afterward to the first three-dimensional models for α-keratin fibers[13,14] and collagen.[15,16] It seemed as if the structure of proteins would yield to stereochemical considerations and parametric equations, as that of nucleic acids already had. The determination of the first protein crystal structures however showed a disconcerting irregularity, which suggested that the principles of protein structure were far more complicated.[17]

After it was established in the 1960s that the structure of a protein was fully specified in its sequence,[18] many attempts were made to deduce one from the other by computation—known as the protein folding problem,[19] but success was very low and experimental methods gained uncontested supremacy in establishing the atomic structure of proteins. Nevertheless, computational methods received increasing attention in the 1980s and there was palpable excitement about new biophysical approaches, such as threading[20–21] and simplified representations of the polypeptide chain,[22,23] which promised decisive progress in deducing protein structure from sequence. The failure to translate these approaches into real-life applications raised questions about their usefulness and prompted the first CASP experiment (Critical Assessment of Structure Prediction), a double-blind test of predictive success.[1] The results were sobering. Predictive methods were found to be quite blunt and the only method that worked reliably was to identify a homolog of known structure to the target sequence and model the target on it (known as template-based modeling, TBM, or as homology modeling). After initial progress in developing ever more sensitive

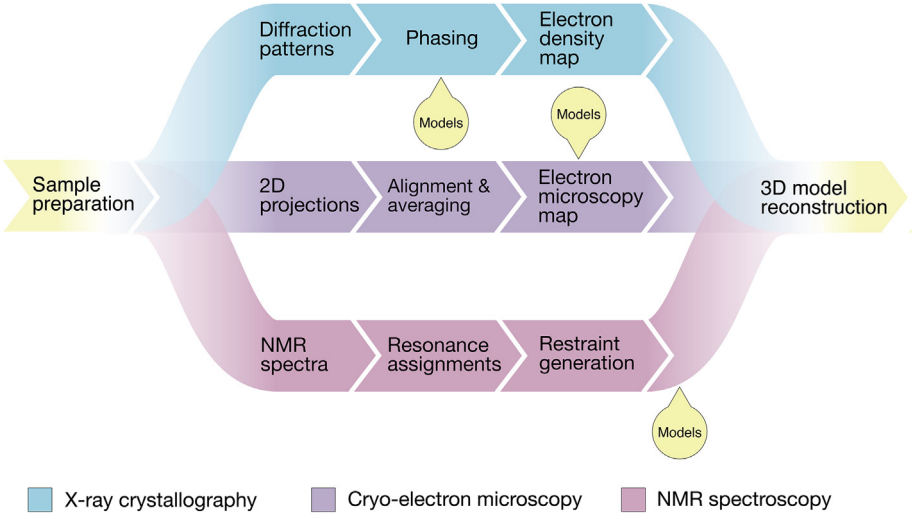

Figure 1. Influence of accurate modeling methods on workflows for experimental structure determination. Steps, where a predicted 3D model can be inserted to improve the workflow, are marked with yellow "balloons".

sequence comparison tools, which gave access to more and more distant homologs, CASP experiments showed that progress in prediction methodology slowed to a virtual standstill for more than a decade around 2010.

A new approach however began developing in this time, which eventually led to the stunning breakthrough of AlphaFold2 at CASP14. This breakthrough relies on the realization that evolutionary information captured in multiple alignments of homologs not only offers the occasional template of known structure, but also a wealth of inter-residue correlations that can be translated into contact maps. These in turn can serve as restraints for structure determination, as in NMR experiments. Two important advances had to be achieved first, though: distinguish between direct (structural) and indirect (functional) correlations by introducing direct coupling analysis[24–26] and derive distance maps from residue correlations by deep learning methods.[27] These innovations led to substantial progress from CASP12 (2016) to CASP13 (2018) and were translated into a further decisive advance by AlphaFold2 in CASP14 (2020), by using end-to-end training on a deep-learning network that combined attention modules to derive distance constraints with 3D equivariant transformer neural networks to obtain structural models. Although AlphaFold2 clearly benefited from superb software engineering, it is not unreasonable to expect that other methods will make substantial progress and maybe even catch up, having knowledge of the strategy that led AlphaFold2 to such impressive success. It therefore seems likely that within a few years, the life science community will see a number of highly performant and publicly available structure prediction servers or databases of pre-computed models, which will make the protein structure space as accessible as BLAST made the sequence space 25 years ago.

Given this unforeseen advance, an important question for the structural biology community is how our understanding of protein structure will be impacted. This special issue is dedicated to this question. In our article, we want to look more specifically at how established experimental methods could be empowered by this advance. For this, it is important to keep in mind that, while the protein folding problem has sometimes been phrased narrowly as the effort to deduce the three-dimensional structure of a protein from its sequence, there is far more information encoded in the sequence of a protein, including its dynamics, its folding pathway, recognition of its interaction partners, and its response to changing environments. Access to this information is understood to require the combination of different methods, experimental and computational, in what has been variously referred to as hybrid or integrative approaches. Examples of this are the combination of mass spectrometry and cryo-electron microscopy with molecular modeling to reconstruct the molecular architectures of macromolecular complexes,[28] or the combination of crystallography and NMR with Bayesian inference to determine the structure of the voltage-dependent anion channel.[29] With a new generation of computational structure prediction tools, it is likely that decisive progress will be possible on many other questions in protein structure, which are of the highest biological relevance.

X-ray crystallography

X-ray crystallography has been the most common method to determine the molecular and atomic structure of macromolecules.[30–32] It requires a crystalline form that diffracts a beam of X-rays,[33] which is captured by specialized detectors that visualize the resulting diffraction pattern, or “reflections”. By measuring the angles and intensities of the generated reflections, recorded as the crystal is gradually rotated, it is possible to resolve a three-dimensional picture of the density of electrons within the crystal. From this electron density, the atom coordinates can be determined using mathematical tools.

One critical obstacle for structure determination using X-ray diffraction data is the “phase problem”, which prevents direct determination of the 3D structure.[34] The phase problem extends beyond X-ray diffraction to electron diffraction, which is also used to determine macromolecular structures.[35–37] There are several established ways to solve the crystallographic phase problem, which is also referred to as “phasing”. These ways include direct (ab initio) methods,[38–40] which require high-resolution diffraction data (better than 1.2 Å), isomorphous replacement,[41–44] Single- or Multi-wavelength Anomalous Dispersion (SAD or MAD, respectively) methods,[45,46] which require non-trivial chemical modifications or additions to the macromolecule, or the molecular replacement (MR) method, which uses phases derived from a “template” structure, often a homology model.[47–49] The well-established methods of model assessment and refinement in X-ray crystallography[50–52] provide a reliable approach to using such templates, even those that do not completely recapitulate the determined structure. Structural genomics projects have promoted high-throughput studies and automation at different levels ranging from target selection to structure determination, including the search of initial phases and model building.[53–56] In addition, phasing using native sulfurs present in proteins has also been implemented.[56,57] While most structures solved to date by X-ray crystallography utilized MR with previously solved experimental structures, in some cases such a solution was hampered by a lack of suitable structural templates. Protein structure prediction methods offer new avenues for model building and

refinement,[58–61] including phasing with protein fragment models[62] instead of models of entire proteins. Accordingly, improved and novel models generated by accurate and reliable methods can facilitate crystal structure determination,[4,63,64] especially if such future tools will provide not a single solution, but rather an array of potential conformational states that will increase the chances of capturing a particular crystallized state. The latter is particularly relevant for proteins that can adopt more than one conformational state,[65] which can depend on factors such as crystallization conditions, ligands, or protein partners. Similarly, successful structure prediction of self-assembling proteins and peptides forming polymorphic supramolecular fibrils would facilitate crystal structure determination of these challenging systems, which often lack a suitable model template.[66–70] A timely example for the advantageous capabilities of accurate modeling in X-ray crystallography was recently demonstrated by Flower and Hurley, reporting that the crystal structure of the ORF8 protein of SARS-CoV-2, determined experimentally using the SAD phasing method,[71] could have been solved using MR with an AlphaFold2 model as a template (CASP14 model ID: T1064TS427_1-D1).[72]

Nuclear Magnetic Resonance Spectroscopy (NMR)

NMR is a second widespread method to obtain information about the structure and dynamics of proteins.[73] It is actually a collection of experimental strategies designed to measure the chemical shifts of isotopically labeled nuclei upon electromagnetic stimulation in a powerful magnetic field and the transfer of the magnetization to other nuclei. This transfer can either occur through covalent bonds, which can be used to assign chemical shifts to individual nuclei, or through space, irrespective of the bonded structure, used to generate distance restraints for the subsequent structure calculation.

An important problem in NMR experiments is the overlap between peaks in the spectra, which increases with the size of the protein. This is due to different nuclei in the protein having the same or very similar chemical shifts. While this problem can be alleviated by multidimensional experiments, in which the protein is labeled with two or more different isotopes, size causes a second problem that is harder to address—the magnetization relaxes faster, making peaks broader and weaker. Thus, although NMR studies of megadalton complexes of known structure have been successfully performed through the selective isotope labeling of specific residues,[74] the upper size limit for the *de novo* structure determination of proteins by NMR is largely in the range of single-domain proteins.

This suggests that among protein structure determination methods, NMR may be impacted the most by the new generation of structure prediction methods. These produce their best results on single-domain proteins, use mainly distance restraints similar to NMR, but work orders of magnitude faster. Because structures determined by NMR do not have associated measures of model quality, such as R-factors or the resolution of the underlying data, it will be difficult to assess the relative quality of the models generated by experiment versus computation. Thus, it will be challenging to justify the investment of months of experimental work *versus* a few hours of computer time.

This however neglects one major strength of NMR experiments, namely access to protein dynamics.[75] Dynamics are an essential aspect of protein function and the ability of computational methods to cover them is at present uncertain at best. The combination of computation and experiment would offer a direct path to determining the multiple conformational states a protein can assume in solution. This could be achieved by comparing experimental NMR spectra with expectation spectra back-calculated across a modeled conformational space.[76] The comparison of computational spectra to experimental ones would inherently deliver an R-factor for structure calculation and thus a measure of model quality, and the availability of high-quality computational models for the back-calculation step would speed up the structure calculation enormously.[77] Therefore, the study of protein dynamics by NMR might be greatly empowered by the combination with new structure prediction methods.

Cryo-electron microscopy (cryo-EM)

The third main method for experimental structure determination is cryo-EM, which determines structures of macromolecules in a flash-frozen solution[78,79] or directly in the cell.[80–82] It has become a widely used method to solve structures of macromolecular complexes and, increasingly, small monomeric proteins.[83,84] Structure determination by cryo-EM is a stepwise process.[78] Firstly, thousands of 2D images of the macromolecule of interest are taken. Second, the images, which represent the projections of the molecule from different angles, are aligned and averaged to obtain a 3D EM density map. Lastly, a structural model is built based on the EM map. In most cases, the model building step currently requires starting structures, which are fitted into the map and refined based on the EM density. Such starting structures are usually obtained by X-ray crystallography, NMR, or in some cases - current modeling methods.

Methods for predicting structural models of single proteins with accuracy that equals or exceeds that of AlphaFold2 would definitely make a powerful combination with cryo-EM. Starting models could be built in a matter of hours or days, instead of the months or years usually required for cloning gene

constructs, optimizing purification protocols, and solving structures experimentally. Perhaps less obviously, new methods could provide more complete models than structures solved by X-ray crystallography. Indeed, X-ray structures often have loops or termini missing because of poorly resolved density. Moreover, in crystallography, "troublesome" domains are often cut away to solve problems with purification or crystallization. Finally, AlphaFold2 and future prediction algorithms might preferentially model conformations adopted within a complex, even though these methods predict monomeric models. This is because deep-learning methods rely on sequence alignments of protein families, which also record evolutionary constraints imposed by protein–protein interactions. Thus, new modeling methods might make cryo-EM model building not only quicker but also better.

What could limit the combination of accurate modeling methods and cryo-EM? First, at a resolution better than 3–4 Å, cryo-EM structures can be built without any starting structure using an increasingly growing number of automated programs.[85–88] As cryo-EM advances, such high-resolution data might become routine, even for small proteins. Theoretical modeling methods would not be necessary to solve such cases, but could be useful to either speed up model building by providing a starting model, or to fill parts of the EM map that are at a lower resolution.

Second, at an intermediate resolution (in the range between 3–4 Å and 10 Å), at which only secondary structures and bulky side chains are visible, the accuracy of modeling algorithms might need to be considered explicitly. In this resolution range, models often serve as starting models to be flexibly refined into the EM map.[89–91] By fitting the models into the EM map, it can be relatively easy to assess the accuracy of the models. However, EM densities are not always sufficient to verify whether a predicted model has the correct sequence register along the backbone. Also, loops and linkers that are poorly resolved in the EM density may not match those in the models. Thus, an accurate assessment of the local quality of the models will be necessary to resolve regions where the EM and models do not agree or where the EM density is lacking. While this problem applies also to solving X-ray structures using MR with predicted models, it will be an even more important aspect when using accurate prediction methods in cryo-EM, when some cryo-EM modelers might place too much trust in the local accuracy of the models.

Finally, at a resolution worse than 10 Å, models are usually fitted to the EM maps as rigid bodies, without an additional refinement of their conformations.[89,92,93] The importance of this resolution range cannot be underestimated—maps of low resolution will be provided increasingly in the coming years from cellular cryo-ET.[80–82] Here, again, reliable assessment of model quality, both locally but also globally, will be crucial, as low-resolution EM maps often do not allow to discriminate between alternative structural folds, let alone validate the predicted sequence register. The cryo-EM community will need to create standards on how to annotate and interpret the uncertainty of such models.

Integrative structural biology

Integrative structural biology involves determining structures of macromolecules by combining different experimental and modeling techniques.[94–97] For example, to determine the structure of a protein complex, X-ray crystallography or NMR might be used to solve structures of individual subunits. EM or Small Angle X-ray Scattering (SAXS)[98,99] can be used to obtain the overall shape of a complex.[100,101] In addition, mass spectrometry (MS) techniques coupled with crosslinking[97,100,102–104] or hydrogen–deuterium exchange[105,106] can be applied to obtain spatial restraints on distances or interactions between subunits. A model of the entire complex is then built using integrative modeling methods by fitting individual subunits to the EM map or SAXS data while satisfying restraints from additional techniques.[100,101,103,104,107–109]

The ability to model protein complexes as accurately as monomeric proteins has not been demonstrated yet in unbiased (blind) experiments such as CAPRI (Critical Assessment of PRedicted Interactions).[110] However, a recent report of RoseTTAFold performance shows promise for modeling complexes using deep learning architectures trained on monomeric proteins.[6] Moreover, accurate modeling of single proteins will certainly revolutionize integrative structural biology by easily providing better building blocks for current integrative modeling approaches. All the above-discussed enhancements in experimental methods resulting from new accurate monomeric models will positively impact integrative structural biology. Finally, the accuracy of contact prediction within monomeric proteins, together with promising results from methods predicting inter-protein contacts, such as EVcomplex,[111] suggests that it should also be possible to accurately predict contacts between proteins. Such contacts could be easily adapted as restraints in current integrative modeling software. Indeed, a recent study demonstrated the applicability of this approach to fitting individual subunit structures into low-resolution EM maps.[112] Future methods that can predict contacts at high accuracy and sensitivity can therefore be a game-changer for integrative modeling.

Challenges

Proteins that can adopt multiple conformations

The ability of many proteins to adopt more than one conformation will be a challenge when using

accurate prediction methods in structure determination or prediction pipelines. Multiple conformations are often inherent to function.[9] For example, enzymes and receptors shift between active and inactive conformations, transporters alternate between inward- and outward-facing conformations upon substrate transport, or viral fusion proteins transform between dramatically different pre- and post-fusion conformations. Indeed, it has been long recognized that in the PDB,[113] one can find entries with similar, or even identical sequences, that nevertheless are structurally different from each other.[65,114–117] Structural differences among such sequence-similar entries can far exceed anticipated thermal fluctuations.[118] Known examples of multiple physiologically relevant conformations may serve as templates for modeling missing states of proteins where only a single conformation is available.[119]

In this respect, we posit that the goal of accurate prediction methods should not be to predict a single "best" structure for a query protein. The latter has been an underlying assumption behind CASP assessments, and a single model is often the output of automatic homology modeling servers. Rather, the goal should be to predict the full spectrum of biologically relevant conformations, including those that are dependent on external or additional factors, for example, substrate binding. Since accurate prediction methods can be trained over the entire PDB, including sequence-redundant but structurally dissimilar entries, they can presumably output multiple conformations. Also, information about functional conformations is likely to be present in the multiple sequence alignments used for modeling. Thus, there is reason to hope that future methods will be able to predict all biologically relevant conformations for a query protein. In fact, it may well be that AlphaFold2 or similar methods already have this capacity.

This hope is supported by AlphaFold2's performance in CASP14. AlphaFold2 predicted the structures of 75 CASP14 targets, submitting five different models for each target. We manually examined these models to determine the extent to which they correspond to multiple conformations. Specifically, for each CASP14 target predicted by AlphaFold2, all five models were superimposed and visually inspected. If significant differences between two or more models were observed, such as movement of a helical bundle, we assumed the prediction pointed to multiple conformations. Models that differed only in N- or C-termini devoid of secondary structures were assumed to represent the same conformation. Of these 75 targets, the five different models for 60 targets (80%) showed virtually the same conformation (neglecting differences akin to thermal fluctuations). On the other hand, fifteen targets (20%) manifested more than a single distinct conformation. It is noteworthy, however, that because occasionally the shift between conformations is minor, for example between the active and inactive conformations of an enzyme, we allowed ourselves to be liberal in what we considered here different conformations. In this respect, fifteen targets should be regarded as an upper bound. We discuss two interesting cases, where the predicted multiple conformations could be directly related to biological function.

These two CASP14 targets (T1024 and T1098) belong to the Major-Facilitator Superfamily (MFS). This superfamily of secondary active transporters comprises 76 different families that transport an array of different compounds across the membrane, ranging from ions and amino acids to peptides and oligosaccharides.[120] They operate via an alternating-access mechanism, in which the binding site is alternately accessible from either side of the membrane. This is facilitated by the relative movement of two discrete helical-bundles.[120] It follows that MFS transporters have at least two extreme conformations that are easy to distinguish—inward- and outward-facing (IF and OF, respectively). Indeed, for targets T1024 and T1098 AlphaFold2 predicted both conformations (Figure 2).

Target T1024 represents the multidrug transporter LmrP from *Lactococcus lactis* that recently had its structure solved in a ligand-bound OF conformation (PDB entry 6T1Z).[121] AlphaFold2 predicted one semi-occluded OF conformation (model 3), one occluded-OF conformation (model 5), and three fully open IF conformations (models 1, 2, and 4) of LmrP. Model 3, featuring a semi-occluded OF conformation, was the closest to LmrP's crystal structure with an RMSD of 2.0 Å when superimposing all alpha carbons (Figure 2(A and C)). As for the IF models, these resemble the IF conformations of evolutionary remote MFS members such as the peptide transporter YbgH from *Escherichia coli* (PDB entry 4Q65; Figure 2(B and D)). In this respect, it is noteworthy that in a Perspective, published while this article was in review, Mchaourab and co-workers correlated the predicted IF conformation of this transporter with their own EPR DEER spectroscopy data, providing direct validation of this new conformation.[122]

Target T1098 belongs to the sucrose transport protein SUT1 from *Oryza sativa* (UniProt entry Q10R54), for which an experimental structure is yet to be published. AlphaFold2 predicted two IF (models 1–2) and three OF models (models 3–4; Figure 2(E and F)). With the closest structural homologs available presenting a sequence identity of no more than 16%, it is difficult to assess the overall quality of both the IF and OF models. Although this target was later canceled due to an unverified structure, it does not affect

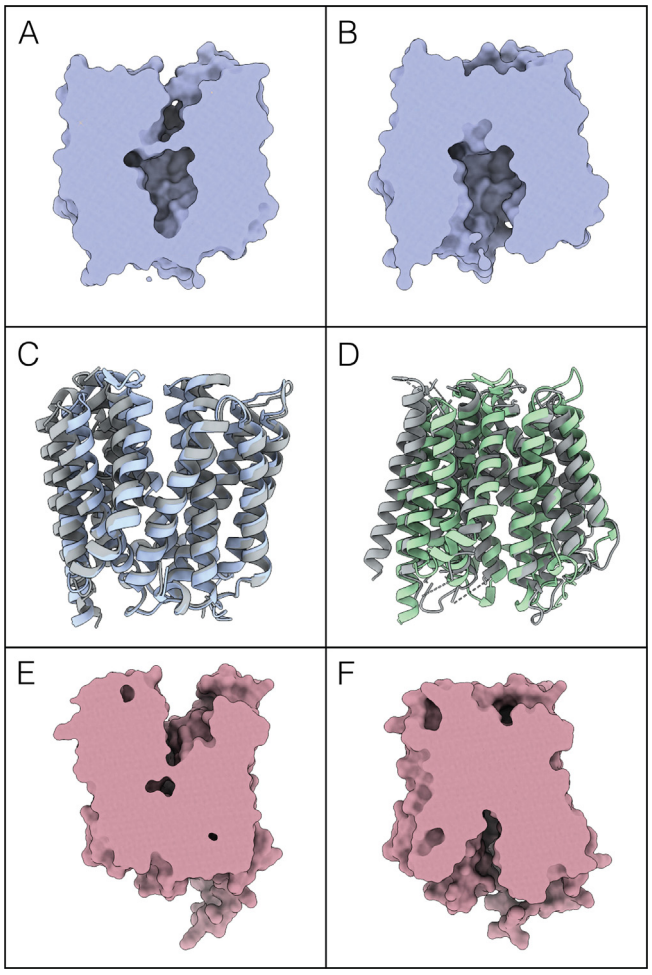

Figure 2. CASP14 targets T1024 and T1098. (A and B) Cross-section of *Escherichia coli* multidrug transporter LmrP (T1024) models by AlphaFold2, in OF (A) and IF (B) conformations. The periplasmic side is at the top and the cytoplasmic side at the bottom. (C) Superimposition of one of AlphaFold2's OF models (blue) and the LmrP crystal structure (gray; PDB entry 6T1Z). The α-carbon RMSD between the two structures is 2.0 Å. (D) Superimposition of one of AlphaFold2's IF models (green) and the peptide transporter YbgH from Escherichia coli (PDB entry 4Q65) suggested by the HMM-based homology detection algorithm HHpred as a potential template. YbgH and LmrP have a sequence identity of 13%. (E and F) Cross-section of *Oryza sativa* sucrose transport protein SUT1 (T1098) models by AlphaFold2 in an OF (A) and IF (B) conformations.

AlphaFold2's predictions that clearly show AlphaFold2 can output multiple biologically relevant conformations for these proteins.

In light of the relatively low sequence identity to previously known templates, AlphaFold2's achievements are impressive, especially when considering that the second-best performing algorithm (Zhang-TBM) predicted only one of these conformations (OF). The 2.0 Å α-carbons RMSD between the experimental structure of LmrP and AlphaFold2's model demonstrate that the method performed significantly better compared to standard homology modeling based on such remote templates. It should be noted, however, that templates for both conformations of LmrP and SUT1 are relatively abundant in the PDB. Indeed, the HMM-based homology detection algorithm HHpred[123] identifies more than 30 different templates for each target, with good coverage and sequence identity ranging between 10% and 16%. Because it is reasonable to assume that AlphaFold2 included these templates in its training set, it is unclear whether it can predict conformations of query proteins with architectures beyond the ones in the PDB. This is a critical question to address when developing future accurate prediction methods. Currently, it is difficult to assess AlphaFold2's true potential for predicting multiple biologically relevant conformations for a wide variety of proteins. This is due to the limited number of CASP targets, and the inherent bias resulting from participants submitting only models that are the most likely to win the competition, rather than represent possible biological diversity. That being

said, the capability to predict such multiple conformations, demonstrated by AlphaFold2 in CASP14, is by itself an important step forward in the field of protein structure prediction. In this respect, it is noteworthy that multiple conformations, regardless of their biological relevance, can serve as better building blocks than single conformations for EM, and as templates in molecular replacement for X-ray and NMR structure determination when the target can in fact adopt multiple conformations.

Macromolecular complexes and self-assembling protein fibrils

One of the challenges of future prediction methods will be the accurate prediction of supra-molecular structures. An interesting example is provided by the outer-membrane "bushing" of the *Salmonella enterica* flagellar basal body (PDB 7BGL),[124] formed by stacked 26-mer rings of FlgI subunits and FlgH-YecR heterodimers, respectively.[124] The stacked rings of FlgH and FlgI were entered into CASP14 as target T1047 and were divided by domains into targets T1047s1-D1 (FlgH) and T1047s2-D1 to -D3 (FlgI). FlgH anchors the rings in the outer membrane via a very long β-hairpin extension, which associates into a 52-stranded β-barrel in the membrane. This proved to be the hardest target for the prediction community. While AlphaFold2 submitted the best model, it was nevertheless its worst prediction, with a GDT_TS value barely above 50. Two effects appear to have led to this poor outcome, both related to the supra-molecular assembly. First, without the realization that the long hairpin associated into a β-barrel in the membrane, it was modeled in a curled-up conformation, rather than extended as in the native structure. Second, the N-terminal extension of the protein, which lacks regular secondary structure and is domain-swapped to the next subunit in the ring, was modeled correctly, but into the binding site of its own subunit. Clearly, the prediction failed to distinguish between inter-chain and intra-chain restraints.[3] The FlgI ring was not nearly as challenging, but the C-terminal domain of the protein (T1047s2-D3), which is also intimately involved in supra-molecular assembly, was the target on which AlphaFold2 posted the worst relative performance, ranked 78th of 117 groups. Each FlgI domain contributes a β-hairpin to complete a β-sheet in the next subunit and AlphaFold2 modeled the hairpin, which is largely unresolved in the cryo-EM structure and presumably interacts with the next ring of subunits in the native basal-body complex, as a four-stranded β-sheet. The top-ranking models in this case were mostly built on the prediction of the tFold server (https://drug.ai.tencent.com/console/en/tfold), from the AI unit of the Chinese technology company Tencent. We conclude that inter-chain contacts in supra-molecular assemblies pose a major challenge to deep-learning methods, but that some methods are already further along in accounting for them than others.

Another difficult area for prediction methods concerns self-assembling targets (Figure 3). For example, different organisms secrete proteins and peptides that self-assemble into ordered amyloid fibrils, which carry various physiological and pathophysiological roles.[125–141] The most studied example is amyloid aggregation associated with fatal neurodegenerative and systemic diseases, which were mostly known to share cross-β structural feature, composed of tightly mated β-sheets.[142–145] Yet, amyloids also function as key virulence factors in microbes,[130,134–137] which has rendered them attractive candidates for structural characterization aimed at discovering novel antimicrobial therapeutics.[66] Amyloid structures have been studied using X-ray crystallography[66–70,146] (cross-β steric-zipper spine segment structures resolved mostly by Eisenberg and co-workers are reviewed by Nelson and Eisenberg),[147] NMR,[148–153] micro-electron diffraction,[154] and cryo-EM.[155–187] Structures to date revealed vast polymorphisms in fibril morphology and arrangement.[132,152,156,188–194] Especially profound is the polymorphism of secondary structure found for the same or closely homologous sequences[66,69,181,195] (Figure 3). Therefore, amyloids and other functional fibrils also challenge the classic "one sequence - one structure - one function" paradigm suggested by the work of Mirsky, Pauling, Anfinsen, and others.[18,196–198] The ability of 3D prediction methods to predict such structural polymorphism is a challenge that remains to be adequately addressed.[199]

Moreover, the inter-molecular interfaces in self-assembling fibrils are tighter compared to other quaternary structures and are often "dry", i.e., lacking water molecules between protein chains, thus presenting a unique type of protein–protein interfaces.[145] Such structures might differ substantially from the training sets used for deep learning, and suggest that special attention should be devoted to accurate prediction methods for such proteins. Amyloid fibrils, in particular, show little conservation in sequence and length, limiting the size of datasets and alignments that can be used to train deep learning algorithms. These unique properties of amyloids and other functional fibrils are expected to present greater challenges to future prediction methods that rely on training sets of known structures, and their success in this area remains to be determined.

Perspectives

While AlphaFold2 and its successors might not render experimental structure determination obsolete, they can certainly redefine X-ray crystallography and NMR. Seeing how multiple prediction methods increased their accuracy between CASP13 and CASP14, and witnessing

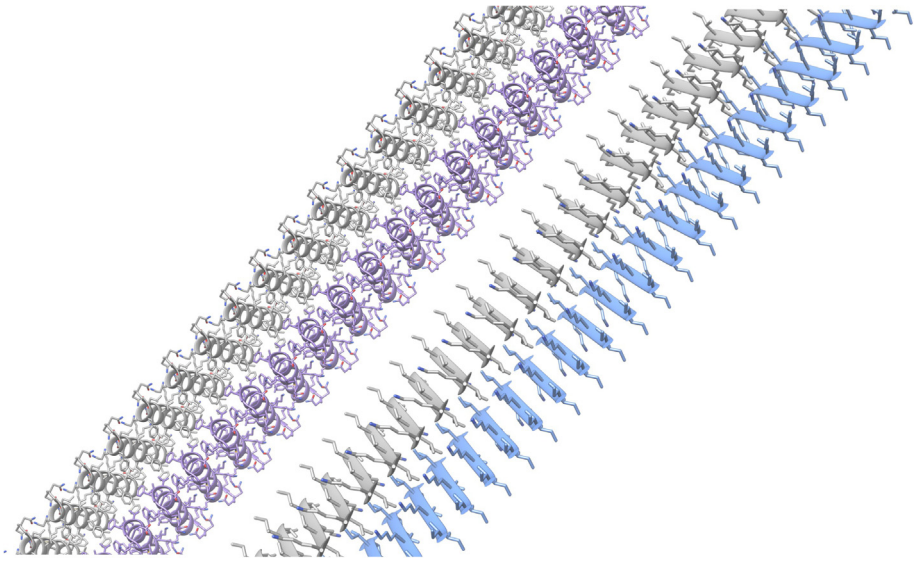

Figure 3. The cross-α crystal structures PSMα3[70] (left, grey and purple ribbons) and the cross-β structure of the IIKVIK amyloid spine segment from PSMα1 (right, grey and blue ribbons),[69] members of the *Staphylococcus aureus* phenol-soluble modulin (PSM) virulent peptide family. Sidechains are shown as sticks and colored by atom type. Both structures encompass molecules stacked perpendicular to the fibril axis, forming mated sheets, yet demonstrate secondary structure polymorphism with PSMα1-IIKVIK forming the canonical amyloid cross-β structure of mated β-sheets, and PSMα3 forming a fibril composed of α-helices stacked in two mated "sheets".

the recent release of the RoseTTAFold program[6] and the preview of the release of AlphaFold2,[5] it is reasonable to expect the emergence of widely and freely available methods for accurate structure prediction. Given such methods, X-ray crystallography and NMR might not be the first to try when characterizing the general structure of a protein. Rather, these experimental methods might focus on studying alternative functional conformations, solving structures in complexes with ligands, studying enzymatic reactions at very high resolution, visualizing the rearrangement of hydrogen bonds and bound water molecules, or screening for drug molecules.

In contrast, at first glance it seems that new modeling methods will empower cryo-EM in its current form by providing accurate building blocks for model reconstruction. However, it is not unreasonable to expect new methods that will also accurately model macromolecular complexes. Such methods could redefine the purpose of single-particle cryo-EM as well. Cryo-EM would then focus on determining the functional states of protein complexes or identifying alternative quaternary conformations, similarly to X-ray crystallography and NMR for smaller or monomeric proteins. We believe this is a promising vision—crystallographers, NMR spectroscopists, and cryo-electron microscopists would have more resources to focus on answering biological questions, instead of fiddling with the orientations of amino acid side chains.

A point to consider is the ability of the general life-science community to make use of accurate models becoming available for most proteins. Even today, many scientists do not fully use structural resources that already exist, such as deposited experimental structures or the analysis tools found in the three PDB websites—at least in part because of insufficient training. Collaborations with structural biologists can alleviate this gap, but this might not be a viable option if accurate prediction methods become the first choice for many scientists. Therefore, incorporating more advanced structure analysis into undergraduate and graduate student training should be prioritized sooner rather than later.

Is an "integrative AlphaFold" possible for one-step automated model building by integrating structures from X-ray crystallography, EM, and other experimental data such as cross-linking mass spectrometry? Currently, this goal seems far-reaching for supervised machine learning methods such as AlphaFold2 because of limited training data, the diverse types of experimental data used in integrative modeling, and differences between datasets of similar types resulting from different instruments and data processing software. Nevertheless, future predictive approaches that do not require massive labeled training data sets, such as Reinforcement Learning,[200] might be applied to optimization problems in integrative structural modeling. Also, deep learning approaches are already used to pre-process data used in integrative modeling, for example, to reduce noise in the data[201] or detect and classify particles in EM micrographs[202] or cryo-ET tomograms of cells.[203]

Finally, is it time to reinvent model building based on X-ray crystallography and EM data? Besides AlphaFold2, most structure prediction methods are composed of complex pipelines. Such pipelines often involve steps such as splitting sequences into predicted protein domains, predicting secondary structures and residue contacts, and performing multi-step modeling protocols. AlphaFold2 apparently replaced such pipelines with the so-called end-to-end approach.[204,205] In this approach, AlphaFold2 takes a sequence as input and returns a 3D structure as output, learning all the steps "behind the scenes". Structure determination using X-ray crystallography and cryo-EM also involve complex pipelines, which start with 2D diffraction patterns or EM images, then reconstruct 3D densities and build 3D models based on those densities. Yet, a vast amount of training data in the form of diffraction patterns and EM images is available. Thus, it might be possible to train deep learning systems that replace current X-ray crystallography and EM model building pipelines with a single step.

CRediT authorship contribution statement

Gal Masarati: Data curation, Visualization, Investigation, Writing - original draft, Writing - review & editing. **Meytal Landau:** Conceptualization, Visualization, Writing - original draft, Writing - review & editing. **Nir Ben-Tal:** Conceptualization, Writing - original draft, Writing - review & editing. **Andrei Lupas:** Conceptualization, Writing - original draft, Writing - review & editing. **Mickey Kosloff:** Conceptualization, Writing - original draft, Writing - review & editing. **Jan Kosinski:** Conceptualization, Writing - original draft, Writing - review & editing.

Acknowledgments

NB-T acknowledges the support of Grant 450/16 of the Israeli Science Foundation (ISF) and Abraham E. Kazan Chair in Structural Biology, Tel Aviv University. AL was supported by institutional funds from the Max Planck Society. MK was supported by funding from the Canadian Institutes of Health Research (CIHR), the International Development Research Centre (IDRC), the Israel Science Foundation (ISF), and the Azrieli Foundation (grant number 3512/19), the NSF-BSF (grant number 2020627), and the DS Research Center at the University of Haifa. JK was supported by funding from the Federal Ministry of Education and Research of Germany (FKZ 031L0100).

Author contributions

G.M. and M.L. prepared the figures, G.M. analyzed the data, G.M., M.L., N. B-T., A.L, M.K., and J.K. wrote the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received 17 May 2021;
Accepted 28 June 2021;
Available online 3 July 2021

Keywords:
AlphaFold2;
protein structure prediction;
protein conformations;
X-ray crystallography;
NMR;
cryo-electron microscopy

References

1. Moult, J., Pedersen, J.T., Judson, R., Fidelis, K., (1995). A large-scale experiment to assess protein structure prediction methods. *Proteins: Struct. Funct. Bioinf.*, **23**, ii–iv.
2. Pearce, R., Zhang, Y., (2021). Deep learning techniques have significantly impacted protein structure prediction and protein design. *Curr. Opin. Struct. Biol.*, **68**, 194–207.
3. Kinch, L., Pei, J., Kryshtafovych, A., Schaeffer, R., Grishin, N., (2021). Topology Evaluation of Difficult Targets in the 14th Round of the Critical Assessment of Protein Structure Prediction (CASP14). *Proteins: Struct. Funct. Bioinf.*
4. Pereira, J., Simpkin, A.J., Hartmann, M.D., Rigden, D.J., Keegan, R.M., Lupas, A.N., (2021). High-accuracy protein structure prediction in CASP14. *Proteins: Struct. Funct. Bioinf.*
5. Demis Hassabis, Brief update on some exciting progress on #AlphaFold! We've been heads down working flat out on our full methods paper (currently under review) with accompanying open source code and on providing broad free access to AlphaFold for the scientific community. More very soon! https://t.co/uP7uzgGMSf, @demishassabis. (2021). https://twitter.com/demishassabis/status/1405922961710854144 (accessed June 23, 2021).
6. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., et al., (2021). Accurate prediction of protein structures and interactions using a 3-track network. *BioRxiv*,. https://doi.org/10.1101/2021.06.14.448402.
7. Xu, D., Zhang, Y., (2012). Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Struct. Funct. Bioinf.*, **80**, 1715–1735.

8. Roy, A., Kucukural, A., Zhang, Y., (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protoc.*, **5**, 725–738.
9. Kessel, A., Ben-Tal, N., (2018). Introduction to Proteins: Structure, Function, and Motion. CRC Press.
10. Astbury, W.T., Street, A., (1931). X-ray studies of the structure of hair, wool, and related fibres.-i. General. *Philos. Trans. R. Soc. Lond. Ser. A, Contain. Pap. Math. Phys. Charact.*, **230**, 75–101.
11. Pauling, L., Corey, R.B., (1950). Two hydrogen-bonded spiral configurations of the polypeptide chain. *J. Am. Chem. Soc.*, **72** 5349–5349.
12. Pauling, L., Corey, R.B., (1951). The pleated sheet, a new layer configuration of polypeptide chains. *PNAS*, **37**, 251.
13. Pauling, L., Corey, R.B., (1953). Compound helical configurations of polypeptide chains: structure of proteins of the α-keratin type. *Nature*, **171**, 59–61.
14. Crick, F., (1953). The packing of α-helices: simple coiled-coils. *Acta Crystallogr. A*, **6**, 689–697.
15. Ramachandran, G., Kartha, G., (1954). Structure of collagen. *Nature*, **174**, 269–270.
16. A. Rich, F. Crick, The structure of collagen, 175 (1955) 915–916.
17. Kendrew, J.C., (1961). The three-dimensional structure of a protein molecule. *Sci. Am.*, **205**, 96–111.
18. Anfinsen, C.B., (1973). Principles that govern the folding of protein chains. *Science*, **181**, 223–230.
19. Dill, K.A., Ozkan, S.B., Shell, M.S., Weikl, T.R., (2008). The protein folding problem. *Annu. Rev. Biophys.*, **37**, 289–316.
20. Bowie, J.U., Luthy, R., Eisenberg, D., (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, **253**, 164–170.
21. Jones, D.T., Taylort, W., Thornton, J.M., (1992). A new approach to protein fold recognition. *Nature*, **358**, 86–89.
22. Dill, K.A., (1999). Polymer principles and protein folding. *Protein Sci.*, **8**, 1166–1180.
23. Srinivasan, R., Rose, G.D., (1995). LINUS: a hierarchic procedure to predict the fold of a protein. *Proteins: Struct. Funct. Bioinf.*, **22**, 81–99.
24. Weigt, M., White, R.A., Szurmant, H., Hoch, J.A., Hwa, T., (2009). Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci.*, **106**, 67–72.
25. Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., Zecchina, R., Onuchic, J.N., et al., (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *PNAS*, **108**, E1293–E1301.
26. Marks, D.S., Colwell, L.J., Sheridan, R., Hopf, T.A., Pagnani, A., Zecchina, R., Sander, C., (2011). Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE*, **6**, e28766
27. Wang, S., Sun, S., Li, Z., Zhang, R., Xu, J., (2017). Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput. Biol.*, **13**, e1005324
28. Robinson, C.V., Sali, A., Baumeister, W., (2007). The molecular sociology of the cell. *Nature*, **450**, 973–982.
29. Bayrhuber, M., Meins, T., Habeck, M., Becker, S., Giller, K., Villinger, S., Vonrhein, C., Griesinger, C., et al., (2008). Structure of the human voltage-dependent anion channel. *Proc. Natl. Acad. Sci.*, **105**, 15370–15375.
30. Bragg, W.L., (1912). The specular reflection of X-rays. *Nature*, **90** 410–410.
31. Friedrich, W., Knipping, P., Laue, M., (1913). Interferenzerscheinungen bei roentgenstrahlen. *Ann. Phys.*, **346**, 971–988.
32. Dobson, C.M., (2019). Biophysical techniques in structural biology. *Annu. Rev. Biochem.*, **88**, 25–33.
33. Einstein, A., (1905). Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Ann. Phys.*, **322**, 132–148.
34. Taylor, G., (2003). The phase problem. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 1881–1890.
35. Gemmi, M., Mugnaioli, E., Gorelik, T.E., Kolb, U., Palatinus, L., Boullay, P., Hovmöller, S., Abrahams, J.P., (2019). 3D electron diffraction: the nanocrystallography revolution. *ACS Cent. Sci.*, **5**, 1315–1329.
36. Shi, D., Nannenga, B.L., Iadanza, M.G., Gonen, T., (2013). Three-dimensional electron crystallography of protein microcrystals. *ELife*, **2**, e01345
37. Glaeser, R.M., (1999). Electron crystallography: present excitement, a nod to the past, anticipating the future. *J. Struct. Biol.*, **128**, 3–14.
38. Rogers, D., Wilson, A.J.C., (1955). Solution of the phase problem. I. The centrosymmetric crystal by H. Hauptman and J. Karle. *Acta Crystallogr. A*, **8**, 365–366.
39. McCoy, A.J., Oeffner, R.D., Wrobel, A.G., Ojala, J.R., Tryggvason, K., Lohkamp, B., Read, R.J., (2017). Ab initio solution of macromolecular crystal structures without direct methods. *PNAS*, **114**, 3637–3641.
40. Schneider, T.R., Sheldrick, G.M., (2002). Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1772–1779.
41. Robertson, J.M., (1937). X-ray analysis and application of Fourier series methods to molecular structures. *Rep. Prog. Phys.*, **4**, 332–367.
42. Green, D.W., Ingram, V.M., Perutz, M.F., Bragg, W.L., (1954). The structure of haemoglobin – IV. Sign determination by the isomorphous replacement method. *Proc. R. Soc. Lond. Ser. A. Math. Phys. Sci.*, **225**, 287–307.
43. Cork, J., (1927). The crystal structure of some of the alums. *London, Edinburgh, Dublin Philos Mag. J. Sci.*, **4**, 688–698.
44. Perutz, M., (1956). Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Crystallogr. A*, **9**, 867–873.
45. Bijvoet, J.M., (1954). Structure of optically active compounds in the solid state. *Nature*, **173**, 888–891.
46. Hendrickson, W.A., (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51–58.
47. Evans, P., McCoy, A., (2008). An introduction to molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.*, **64**, 1–10.
48. Read, R.J., (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. Sect. D, Biol. Crystallogr.*, **57**, 1373–1382.
49. Rossmann, M.G., Blow, D.M., IUCr, (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.*, **15**, 24–31.
50. Murshudov, G.N., Vagin, A.A., Dodson, E.J., (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.

51. Sheldrick, G.M., Schneider, T.R., (1997). SHELXL: high-resolution refinement. *Methods Enzymol.*, **277**, 319–343.
52. 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). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.*, **64**, 61–69.
53. Blundell, T.L., Patel, S., (2004). High-throughput X-ray crystallography for drug discovery. *Curr. Opin. Pharmacol.*, **4**, 490–496.
54. Terwilliger, T.C., (2000). Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 965–972.
55. Terwilliger, T.C., Berendzen, J., (1999). Automated MAD and MIR structure solution. *Acta Crystallogr. Sect. D, Biol. Crystallogr.*, **55**, 849–861.
56. Adams, M.W., Dailey, H.A., DeLucas, L.J., Luo, M., Prestegard, J.H., Rose, J.P., Wang, B.-C., (2003). The southeast collaboratory for structural genomics: a high-throughput gene to structure factory. *Acc. Chem. Res.*, **36**, 191–198.
57. Liu, Z., Vysotski, E.S., Vysotski, E.S., Chen, C., Rose, J. P., Lee, J., Wang, B., (2000). Structure of the Ca2+-regulated photoprotein obelin at 1.7 Å resolution determined directly from its sulfur substructure. *Protein Sci.*, **9**, 2085–2093.
58. DiMaio, F., (2017). Rosetta structure prediction as a tool for solving difficult molecular replacement problems. *Protein Crystallogr.*,, 455–466.
59. Heo, L., Feig, M., (2020). High-accuracy protein structures by combining machine-learning with physics-based refinement. *Proteins: Struct. Funct. Bioinf.*, **88**, 637–642.
60. McCoy, A.J., Stockwell, D.H., Sammito, M.D., Oeffner, R. D., Hatti, K.S., Croll, T.I., Read, R.J., (2021). Phasertng: directed acyclic graphs for crystallographic phasing. *Acta Crystallogr. Sect. D, Struct. Biol.*, **77**, 1–10.
61. Jin, S., Miller, M.D., Chen, M., Schafer, N.P., Lin, X., Chen, X., Phillips, G.N., Wolynes, P.G., (2020). Molecular-replacement phasing using predicted protein structures from AWSEM-Suite. *IUCrJ*, **7**, 1168–1178.
62. Richards, L.S., Millán, C., Miao, J., Martynowycz, M.W., Sawaya, M.R., Gonen, T., Borges, R.J., Usón, I., Rodriguez, J.A., (2020). Fragment-based determination of a proteinase K structure from MicroED data using ARCIMBOLDO_SHREDDER. *Acta Crystallogr. Sect. D, Struct. Biol.*, **76**, 703–712.
63. Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A.J., Read, R.J., Baker, D., (2007). High-resolution structure prediction and the crystallographic phase problem. *Nature*, **450**, 259–264.
64. Wang, Y., Virtanen, J., Xue, Z., Tesmer, J.J., Zhang, Y., (2016). Using iterative fragment assembly and progressive sequence truncation to facilitate phasing and crystal structure determination of distantly related proteins. *Acta Crystallogr. Sect. D: Struct. Biol.*, **72**, 616–628.
65. Kosloff, M., Kolodny, R., (2008). Sequence-similar, structure-dissimilar protein pairs in the PDB. *Proteins: Struct. Funct. Bioinf.*, **71**, 891–902.
66. Salinas, N., Tayeb-Fligelman, E., Sammito, M.D., Bloch, D., Jelinek, R., Noy, D., Usón, I., Landau, M., (2021). The amphibian antimicrobial peptide uperin 3.5 is a cross-α/cross-β chameleon functional amyloid. *PNAS*,, 118.
67. Tayeb-Fligelman, E., Salinas, N., Tabachnikov, O., Landau, M., (2020). Staphylococcus aureus PSMα3 cross-α fibril polymorphism and determinants of cytotoxicity. *Structure*, **28**, 301–313.
68. Engelberg, Y., Landau, M., (2020). The Human LL-37 (17–29) antimicrobial peptide reveals a functional supramolecular structure. *Nature Commun.*, **11**, 1–10.
69. Salinas, N., Colletier, J.-P., Moshe, A., Landau, M., (2018). Extreme amyloid polymorphism in Staphylococcus aureus virulent PSMα peptides. *Nature Commun.*, **9**, 1–9.
70. Tayeb-Fligelman, E., Tabachnikov, O., Moshe, A., Goldshmidt-Tran, O., Sawaya, M.R., Coquelle, N., Colletier, J.-P., Landau, M., (2017). The cytotoxic Staphylococcus aureus PSMα3 reveals a cross-α amyloid-like fibril. *Science*, **355**, 831–833.
71. Flower, T.G., Buffalo, C.Z., Hooy, R.M., Allaire, M., Ren, X., Hurley, J.H., (2021). Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion protein. *PNAS*,, 118.
72. Flower, T.G., Hurley, J.H., (2021). Crystallographic molecular replacement using an in silico-generated search model of SARS-CoV-2 ORF8. *Protein Sci.*, **30**, 728–734.
73. Wüthrich, K., (1990). Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.*, **265**, 22059–22062.
74. Mainz, A., Religa, T.L., Sprangers, R., Linser, R., Kay, L. E., Reif, B., (2013). NMR spectroscopy of soluble protein complexes at one mega-dalton and beyond. *Angew. Chem. Int. Ed.*, **52**, 8746–8751.
75. Sekhar, A., Kay, L.E., (2019). An NMR view of protein dynamics in health and disease. *Annu. Rev. Biophys.*, **48**, 297–319.
76. ElGamacy, M., Riss, M., Zhu, H., Truffault, V., Coles, M., (2019). Mapping local conformational landscapes of proteins in solution. *Structure*, **27**, 853–865.
77. Hernandez Alvarez, B., Skokowa, J., Coles, M., Mir, P., Nasri, M., Maksymenko, K., Weidmann, L., Rogers, K.W., et al., (2020). Design of novel granulopoietic proteins by topological rescaffolding. *PLoS Biol.*, **18**, e3000919
78. Cheng, Y., Grigorieff, N., Penczek, P.A., Walz, T., (2015). A primer to single-particle cryo-electron microscopy. *Cell*, **161**, 438–449.
79. Nogales, E., Scheres, S.H., (2015). Cryo-EM: a unique tool for the visualization of macromolecular complexity. *Mol. Cell*, **58**, 677–689.
80. Asano, S., Engel, B.D., Baumeister, W., (2016). In situ cryo-electron tomography: a post-reductionist approach to structural biology. *J. Mol. Biol.*, **428**, 332–343.
81. Pfeffer, S., Mahamid, J., (2018). Unravelling molecular complexity in structural cell biology. *Curr. Opin. Struct. Biol.*, **52**, 111–118.
82. Böhning, J., Bharat, T.A., (2021). Towards high-throughput in situ structural biology using electron cryotomography. *Prog. Biophys. Mol. Biol.*, **160**, 97–103.
83. Kühlbrandt, W., (2014). The resolution revolution. *Science*, **343**, 1443–1444.
84. Nygaard, R., Kim, J., Mancia, F., (2020). Cryo-electron microscopy analysis of small membrane proteins. *Curr. Opin. Struct. Biol.*, **64**, 26–33.
85. Yip, K.M., Fischer, N., Paknia, E., Chari, A., Stark, H., (2020). Atomic-resolution protein structure determination by cryo-EM. *Nature*, **587**, 157–161.

86. Chojnowski, G., Sobolev, E., Heuser, P., Lamzin, V.S., (2021). The accuracy of protein models automatically built into cryo-EM maps with ARP/wARP. *Acta Crystallogr. Sect. D: Struct. Biol.*, **77**
87. Wang, R.Y.-R., Kudryashev, M., Li, X., Egelman, E.H., Basler, M., Cheng, Y., Baker, D., DiMaio, F., (2015). De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nature Methods*, **12**, 335–338.
88. Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P.M.G.E., Grigoras, I.T., Malinauskaite, L., Scheres, S.H.W., et al., (2020). Single-particle cryo-EM at atomic resolution. *Nature*, **587**, 152–156.
89. Malhotra, S., Träger, S., Dal Peraro, M., Topf, M., (2019). Modelling structures in cryo-EM maps. *Curr. Opin. Struct. Biol.*, **58**, 105–114.
90. Trabuco, L.G., Villa, E., Mitra, K., Frank, J., Schulten, K., (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*, **16**, 673–683.
91. Lopéz-Blanco, J.R., Chacón, P., (2013). iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.*, **184**, 261–270.
92. Wriggers, W., (2012). Conventions and workflows for using Situs. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 344–351.
93. Cp van Zundert, G., Mjj Bonvin, A., (2015). Fast and sensitive rigid-body fitting into cryo-EM density maps with PowerFit. *AIMS Biophys.*, **2**, 73–87.
94. Rout, M.P., Sali, A., (2019). Principles for integrative structural biology studies. *Cell*, **177**, 1384–1403.
95. Rantos, V., Karius, K., Kosinski, J., (2021). Integrative structural modelling of macromolecular complexes using Assembline. *BioRxiv*,. https://doi.org/10.1101/2021.04.06.438590.
96. Koukos, P., Bonvin, A., (2020). Integrative modelling of biomolecular complexes. *J. Mol. Biol.*, **432**, 2861–2881.
97. Kosinski, J., Mosalaganti, S., von Appen, A., Teimer, R., DiGuilio, A.L., Wan, W., Bui, K.H., Hagen, W.J.H., et al., (2016). Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science*, **352**, 363–365.
98. Graewert, M.A., Svergun, D.I., (2013). Impact and progress in small and wide angle X-ray scattering (SAXS and WAXS). *Curr. Opin. Struct. Biol.*, **23**, 748–754.
99. Gräwert, T.W., Svergun, D.I., (2020). Structural modeling using solution small-angle X-ray scattering (SAXS). *J. Mol. Biol.*, **432**, 3078–3092.
100. Dauden, M.I., Kosinski, J., Kolaj-Robin, O., Desfosses, A., Ori, A., Faux, C., Hoffmann, N.A., Onuma, O.F., et al., (2017). Architecture of the yeast Elongator complex. *EMBO Rep.*, **18**, 264–279.
101. Bernal, I., Börnicke, J., Heidemann, J., Svergun, D., Horstmann, J.A., Erhardt, M., Tuukkanen, A., Uetrecht, C., et al., (2019). Molecular organization of soluble type III secretion system sorting platform complexes. *J. Mol. Biol.*, **431**, 3787–3803.
102. O'Reilly, F.J., Rappsilber, J., (2018). Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nature Struct. Mol. Biol.*, **25**, 1000–1008.
103. O'Reilly, F.J., Xue, L., Graziadei, A., Sinn, L., Lenz, S., Tegunov, D., Blötz, C., Singh, N., et al., (2020). In-cell architecture of an actively transcribing-translating expressome. *Science*, **369**, 554–557.
104. Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B.D., et al., (2018). Integrative structure and functional anatomy of a nuclear pore complex. *Nature*, **555**, 475–482.
105. Konermann, L., Pan, J., Liu, Y.-H., (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.*, **40**, 1224–1234.
106. Lauer, J., Segeletz, S., Cezanne, A., Guaitoli, G., Raimondi, F., Gentzel, M., Alva, V., Habeck, M., et al., (2019). Auto-regulation of Rab5 GEF activity in Rabex5 by allosteric structural changes, catalytic core dynamics and ubiquitin binding. *ELife*, **8**, e46302
107. Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., Sali, A., (2016). FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.*, **44**, W424–W429.
108. Lasker, K., Förster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., et al., (2012). Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *PNAS*, **109**, 1380–1387.
109. Viswanath, S., Bonomi, M., Kim, S.J., Klenchin, V.A., Taylor, K.C., Yabut, K.C., Umbreit, N.T., Van Epps, H.A., et al., (2017). The molecular architecture of the yeast spindle pole body core determined by Bayesian integrative modeling. *Mol. Biol. Cell*, **28**, 3298–3314.
110. Janin, J., Henrick, K., Moult, J., Eyck, L.T., Sternberg, M. J., Vajda, S., Vakser, I., Wodak, S.J., (2003). CAPRI: a critical assessment of predicted interactions. *Proteins: Struct. Funct. Bioinf.*, **52**, 2–9.
111. Hopf, T.A., Schärfe, C.P., Rodrigues, J.P., Green, A.G., Kohlbacher, O., Sander, C., Bonvin, A.M., Marks, D.S., (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. *ELife*, **3**, e03430
112. C.L. McCafferty, D.W. Taylor, E.M. Marcotte, Improving integrative 3D modeling into low-to medium-resolution EM structures with evolutionary couplings, BioRxiv (n.d.) 2021–01. https://doi.org/10.1101/2021.01.14.426447.
113. Berman, H., Henrick, K., Nakamura, H., (2003). Announcing the worldwide protein data bank. *Nature Struct. Mol. Biol.*, **10** 980–980.
114. Gerstein, M., Krebs, W., (1998). A database of macromolecular motions. *Nucleic Acids Res.*, **26**, 4280–4290.
115. Juritz, E.I., Alberti, S.F., Parisi, G.D., (2011). PCDB: a database of protein conformational diversity. *Nucleic Acids Res.*, **39**, D475–D479.
116. Monzon, A.M., Juritz, E., Fornasari, M.S., Parisi, G., (2013). CoDNaS: a database of conformational diversity in the native state of proteins. *Bioinformatics*, **29**, 2512–2514.
117. Li, W., Kinch, L.N., Karplus, P.A., Grishin, N.V., (2015). ChSeq: A database of chameleon sequences. *Protein Sci.*, **24**, 1075–1086.
118. Burra, P.V., Zhang, Y., Godzik, A., Stec, B., (2009). Global distribution of conformational states derived from redundant models in the PDB points to non-uniqueness of the protein structure. *PNAS*, **106**, 10505–10510.
119. Narunsky, A., Nepomnyachiy, S., Ashkenazy, H., Kolodny, R., Ben-Tal, N., (2015). ConTemplate suggests

possible alternative conformations for a query protein of known structure. *Structure*, **23**, 2162–2170.
120. Yan, N., (2013). Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem. Sci.*, **38**, 151–159.
121. Debruycker, V., Hutchin, A., Masureel, M., Ficici, E., Martens, C., Legrand, P., Stein, R.A., Mchaourab, H.S., et al., (2020). An embedded lipid in the multidrug transporter LmrP suggests a mechanism for polyspecificity. *Nature Struct. Mol. Biol.*, **27**, 829–835.
122. Del Alamo, D., Govaerts, C., Mchaourab, H.S., (2021). AlphaFold2 predicts the inward-facing conformation of the multidrug transporter LmrP. *Proteins Struct. Funct. Bioinf.*,.
123. Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., et al., (2018). A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.*, **430**, 2237–2243.
124. Johnson, S., Furlong, E.J., Deme, J.C., Nord, A.L., Caesar, J.J., Chevance, F.F., Berry, R.M., Hughes, K. T., et al., (2021). Molecular structure of the intact bacterial flagellar basal body. *Nature Microbiol.*,, 1–10.
125. Hughes, M.P., Sawaya, M.R., Boyer, D.R., Goldschmidt, L., Rodriguez, J.A., Cascio, D., Chong, L., Gonen, T., et al., (2018). Atomic structures of low-complexity protein segments reveal kinked beta sheets that assemble networks. *Science*, **359**, 698–701.
126. Soragni, A., Yousefi, S., Stoeckle, C., Soriaga, A.B., Sawaya, M.R., Kozlowski, E., Schmid, I., Radonjic-Hoesli, S., et al., (2015). Toxicity of eosinophil MBP is repressed by intracellular crystallization and promoted by extracellular aggregation. *Mol. Cell*, **57**, 1011–1021.
127. Maji, S.K., Perrin, M.H., Sawaya, M.R., Jessberger, S., Vadodaria, K., Rissman, R.A., Singru, P.S., Nilsson, K.P., et al., (2009). Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science*, **325**, 328–332.
128. Berson, J.F., Harper, D.C., Tenza, D., Raposo, G., Marks, M.S., (2001). Pmel17 initiates premelanosome morphogenesis within multivesicular bodies. *Mol. Biol. Cell*, **12**, 3451–3464.
129. Fowler, D.M., Koulov, A.V., Alory-Jost, C., Marks, M.S., Balch, W.E., Kelly, J.W., (2006). Functional amyloid formation within mammalian tissue. *PLoS Biol.*, **4**, e6
130. Syed, A.K., Boles, B.R., (2014). Fold modulating function: bacterial toxins to functional amyloids. *Front. Microbiol.*, **5**, 401.
131. Fowler, D.M., Kelly, J.W., (2012). Functional amyloidogenesis and cytotoxicity-insights into biology and pathology. *PLoS Biol.*, **10**, e1001459
132. Hu, K.N., McGlinchey, R.P., Wickner, R.B., Tycko, R., (2011). Segmental polymorphism in a functional amyloid. *Biophys. J.*, **101**, 2242–2250.
133. Kummer, M.P., Maruyama, H., Huelsmann, C., Baches, S., Weggen, S., Koo, E.H., (2009). Formation of Pmel17 amyloid is regulated by juxtamembrane metalloproteinase cleavage, and the resulting C-terminal fragment is a substrate for gamma-secretase. *J. Biol. Chem.*, **284**, 2296–2306.
134. DePas, W.H., Chapman, M.R., (2012). Microbial manipulation of the amyloid fold. *Res. Microbiol.*, **163**, 592–606.
135. Blanco, L.P., Evans, M.L., Smith, D.R., Badtke, M.P., Chapman, M.R., (2012). Diversity, biogenesis and function of microbial amyloids. *Trends Microbiol.*, **20**, 66–73.
136. Chapman, M.R., Robinson, L.S., Pinkner, J.S., Roth, R., Heuser, J., Hammar, M., Normark, S., Hultgren, S.J., (2002). Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science*, **295**, 851–855.
137. Schwartz, K., Boles, B.R., (2013). Microbial amyloids-functions and interactions within the host. *Curr. Opin. Microbiol.*, **16**, 93–99.
138. Seuring, C., Verasdonck, J., Gath, J., Ghosh, D., Nespovitaya, N., Wälti, M.A., Maji, S.K., Cadalbert, R., et al., (2020). The three-dimensional structure of human β-endorphin amyloid fibrils. *Nature Struct. Mol. Biol.*, **27**, 1178–1184.
139. Luo, F., Gui, X., Zhou, H., Gu, J., Li, Y., Liu, X., Zhao, M., Li, D., Li, X., Liu, C., (2018). Atomic structures of FUS LC domain segments reveal bases for reversible amyloid fibril formation. *Nature Struct. Mol. Biol.*, **25**, 341–346.
140. Guenther, E.L., Cao, Q., Trinh, H., Lu, J., Sawaya, M.R., Cascio, D., Boyer, D.R., Rodriguez, J.A., et al., (2018). Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nature Struct. Mol. Biol.*, **25**, 463–471.
141. Hewetson, A., Do, H.Q., Myers, C., Muthusubramanian, A., Sutton, R.B., Wylie, B.J., Cornwall, G.A., (2017). Functional amyloids in reproduction. *Biomolecules*, **7**, 46.
142. Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B., Blake, C.C., (1997). Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.*, **273**, 729–739.
143. Jahn, T.R., Makin, O.S., Morris, K.L., Marshall, K.E., Tian, P., Sikorski, P., Serpell, L.C., (2010). The common architecture of cross-beta amyloid. *J. Mol. Biol.*, **395**, 717–727.
144. Iadanza, M.G., Jackson, M.P., Hewitt, E.W., Ranson, N. A., Radford, S.E., (2018). A new era for understanding amyloid structures and disease. *Nature Rev. Mol. Cell Biol.*, **19**, 755–773.
145. Eisenberg, D.S., Sawaya, M.R., (2017). Structural studies of amyloid proteins at the molecular level. *Annu. Rev. Biochem.*, **86**, 69–95.
146. Perov, S., Lidor, O., Salinas, N., Golan, N., Tayeb-Fligelman, E., Deshmukh, M., Willbold, D., Landau, M., (2019). Structural insights into Curli CsgA Cross-β fibril architecture inspire repurposing of anti-amyloid compounds as anti-biofilm agents. *PLoS Pathog.*, **15**, e1007978
147. Nelson, R., Eisenberg, D., (2006). Recent atomic models of amyloid fibril structure. *Curr. Opin. Struct. Biol.*, **16**, 260–265.
148. Van Melckebeke, H., Wasmer, C., Lange, A., Ab, E., Loquet, A., Bockmann, A., Meier, B.H., (2010). Atomic-resolution three-dimensional structure of HET-s(218–289) amyloid fibrils by solid-state NMR spectroscopy. *J. Am. Chem. Soc.*, **132**, 13765–13775.
149. Xiao, Y., Ma, B., McElheny, D., Parthasarathy, S., Long, F., Hoshi, M., Nussinov, R., Ishii, Y., (2015). Abeta(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nature Struct. Mol. Biol.*, **22**, 499–505.
150. Colvin, M.T., Silvers, R., Ni, Q.Z., Can, T.V., Sergeyev, I., Rosay, M., Donovan, K.J., Michael, B., et al., (2016). Atomic resolution structure of monomorphic Abeta42 amyloid fibrils. *J. Am. Chem. Soc.*, **138**, 9663–9674.

151. Walti, M.A., Ravotti, F., Arai, H., Glabe, C.G., Wall, J.S., Bockmann, A., Guntert, P., Meier, B.H., et al., (2016). Atomic-resolution structure of a disease-relevant Abeta (1–42) amyloid fibril. *PNAS*, **113**, E4976–E4984.
152. Qiang, W., Yau, W.M., Lu, J.X., Collinge, J., Tycko, R., (2017). Structural variation in amyloid-beta fibrils from Alzheimer's disease clinical subtypes. *Nature*, **541**, 217–221.
153. Murray, D.T., Kato, M., Lin, Y., Thurber, K.R., Hung, I., McKnight, S.L., Tycko, R., (2017). Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell*, **171**, 615–627.
154. Rodriguez, J.A., Ivanova, M.I., Sawaya, M.R., Cascio, D., Reyes, F.E., Shi, D., Sangwan, S., Guenther, E.L., et al., (2015). Structure of the toxic core of alpha-synuclein from invisible crystals. *Nature*, **525**, 486–490.
155. Schmidt, M., Rohou, A., Lasker, K., Yadav, J.K., Schiene-Fischer, C., Fandrich, M., Grigorieff, N., (2015). Peptide dimer structure in an Abeta(1–42) fibril visualized with cryo-EM. *PNAS*, **112**, 11858–11863.
156. Close, W., Neumann, M., Schmidt, A., Hora, M., Annamalai, K., Schmidt, M., Reif, B., Schmidt, V., et al., (2018). Physical basis of amyloid fibril polymorphism. *Nature Commun.*, **9**, 1–7.
157. Iadanza, M.G., Silvers, R., Boardman, J., Smith, H.I., Karamanos, T.K., Debelouchina, G.T., Su, Y., Griffin, R. G., et al., (2018). The structure of a beta2-microglobulin fibril suggests a molecular basis for its amyloid polymorphism. *Nature Commun.*, **9**, 1–10.
158. Fitzpatrick, A.W.P., Falcon, B., He, S., Murzin, A.G., Murshudov, G., Garringer, H.J., Crowther, R.A., Ghetti, B., et al., (2017). Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature*, **547**, 185–190.
159. Guerrero-Ferreira, R., Taylor, N.M., Mona, D., Ringler, P., Lauer, M.E., Riek, R., Britschgi, M., Stahlberg, H., (2018). Cryo-EM structure of alpha-synuclein fibrils. *ELife*, **7**, e36402
160. Radamaker, L., Lin, Y.-H., Annamalai, K., Huhn, S., Hegenbart, U., Schönland, S.O., Fritz, G., Schmidt, M., et al., (2019). Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis. *Nature Commun.*, **10**, 1–8.
161. Li, B., Ge, P., Murray, K.A., Sheth, P., Zhang, M., Nair, G., Sawaya, M.R., Shin, W.S., et al., (2018). Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nature Commun.*, **9**, 1–10.
162. Li, Y., Zhao, C., Luo, F., Liu, Z., Gui, X., Luo, Z., Zhang, X., Li, D., Liu, C., Li, X., (2018). Amyloid fibril structure of α-synuclein determined by cryo-electron microscopy. *Cell Res.*, **28**, 897–903.
163. Ni, X., McGlinchey, R.P., Jiang, J., Lee, J.C., (2019). Structural insights into α-synuclein fibril polymorphism: Effects of Parkinson's disease-related C-terminal truncations. *J. Mol. Biol.*, **431**, 3913–3919.
164. Guerrero-Ferreira, R., Taylor, N.M., Arteni, A.-A., Kumari, P., Mona, D., Ringler, P., Britschgi, M., Lauer, M.E., et al., (2019). Two new polymorphic structures of human full-length alpha-synuclein fibrils solved by cryo-electron microscopy. *ELife*, **8**, e48907
165. Sun, Y., Hou, S., Zhao, K., Long, H., Liu, Z., Gao, J., Zhang, Y., Su, X.D., Li, D., Liu, C., (2020). Cryo-EM structure of full-length α-synuclein amyloid fibril with Parkinson's disease familial A53T mutation. *Cell Res.*, **30**, 360–362.
166. Boyer, D.R., Li, B., Sun, C., Fan, W., Sawaya, M.R., Jiang, L., Eisenberg, D.S., (2019). Structures of fibrils formed by α-synuclein hereditary disease mutant H50Q reveal new polymorphs. *Nature Struct. Mol. Biol.*, **26**, 1044–1052.
167. Boyer, D.R., Li, B., Sun, C., Fan, W., Zhou, K., Hughes, M.P., Sawaya, M.R., Jiang, L., et al., (2020). The α-synuclein hereditary mutation E46K unlocks a more stable, pathogenic fibril structure. *PNAS*, **117**, 3592–3602.
168. Zhao, K., Lim, Y.J., Liu, Z., Long, H., Sun, Y., Hu, J.J., Zhao, C., Tao, Y., et al., (2020). Parkinson's disease-related phosphorylation at Tyr39 rearranges α-synuclein amyloid fibril structure revealed by cryo-EM. *PNAS*, **117**, 20305–20315.
169. Schweighauser, M., Shi, Y., Tarutani, A., Kametani, F., Murzin, A.G., Ghetti, B., Matsubara, T., Tomita, T., et al., (2020). Structures of α-synuclein filaments from multiple system atrophy. *Nature*, **585**, 464–469.
170. Falcon, B., Zhang, W., Murzin, A.G., Murshudov, G., Garringer, H.J., Vidal, R., Crowther, R.A., Ghetti, B., et al., (2018). Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature*, **561**, 137–140.
171. Zhang, W., Falcon, B., Murzin, A.G., Fan, J., Crowther, R. A., Goedert, M., Scheres, S.H., (2019). Heparin-induced tau filaments are polymorphic and differ from those in Alzheimer's and Pick's diseases. *ELife*, **8**, e43584
172. Falcon, B., Zhang, W., Schweighauser, M., Murzin, A.G., Vidal, R., Garringer, H.J., Ghetti, B., Scheres, S.H.W., et al., (2018). Tau filaments from multiple cases of sporadic and inherited Alzheimer's disease adopt a common fold. *Acta Neuropathol.*, **136**, 699–708.
173. Falcon, B., Zivanov, J., Zhang, W., Murzin, A.G., Garringer, H.J., Vidal, R., Crowther, R.A., Newell, K.L., et al., (2019). Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature*, **568**, 420–423.
174. Zhang, W., Tarutani, A., Newell, K.L., Murzin, A.G., Matsubara, T., Falcon, B., Vidal, R., Garringer, H.J., et al., (2020). Novel tau filament fold in corticobasal degeneration. *Nature*, **580**, 283–287.
175. Arakhamia, T., Lee, C.E., Carlomagno, Y., Duong, D.M., Kundinger, S.R., Wang, K., Williams, D., DeTure, M., et al., (2020). Posttranslational Modifications Mediate the Structural Diversity of Tauopathy Strains. *Cell*, **180**, 633–644.
176. Liberta, F., Loerch, S., Rennegarbe, M., Schierhorn, A., Westermark, P., Westermark, G.T., Hazenberg, B.P.C., Grigorieff, N., et al., (2019). Cryo-EM fibril structures from systemic AA amyloidosis reveal the species complementarity of pathological amyloids. *Nature Commun.*, **10**, 1–10.
177. Swuec, P., Lavatelli, F., Tasaki, M., Paissoni, C., Rognoni, P., Maritan, M., Brambilla, F., Milani, P., et al., (2019). Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain AL amyloidosis patient. *Nature Commun.*, **10**, 1–9.
178. Schmidt, M., Wiese, S., Adak, V., Engler, J., Agarwal, S., Fritz, G., Westermark, P., Zacharias, M., et al., (2019). Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. *Nature Commun.*, **10**, 1–9.

179. Kollmer, M., Close, W., Funk, L., Rasmussen, J., Bsoul, A., Schierhorn, A., Schmidt, M., Sigurdson, C.J., et al., (2019). Cryo-EM structure and polymorphism of Aβ amyloid fibrils purified from Alzheimer's brain tissue. *Nature Commun.*, **10**, 1–8.
180. Gremer, L., Schölzel, D., Schenk, C., Reinartz, E., Labahn, J., Ravelli, R.B.G., Tusche, M., Lopez-Iglesias, C., et al., (2017). Fibril structure of amyloid-β(1–42) by cryo-electron microscopy. *Science*, **358**, 116–119.
181. Röder, C., Kupreichyk, T., Gremer, L., Schäfer, L.U., Pothula, K.R., Ravelli, R.B.G., Willbold, D., Hoyer, W., et al., (2020). Cryo-EM structure of islet amyloid polypeptide fibrils reveals similarities with amyloid-β fibrils. *Nature Struct. Mol. Biol.*, **27**, 660–667.
182. Cao, Q., Boyer, D.R., Sawaya, M.R., Ge, P., Eisenberg, D.S., (2020). Cryo-EM structure and inhibitor design of human IAPP (amylin) fibrils. *Nature Struct. Mol. Biol.*, **27**, 653–659.
183. Lu, J., Cao, Q., Hughes, M.P., Sawaya, M.R., Boyer, D. R., Cascio, D., Eisenberg, D.S., (2020). CryoEM structure of the low-complexity domain of hnRNPA2 and its conversion to pathogenic amyloid. *Nature Commun.*, **11**, 1–11.
184. Hervas, R., Rau, M.J., Park, Y., Zhang, W., Murzin, A.G., Fitzpatrick, J.A.J., Scheres, S.H.W., Si, K., (2020). Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in Drosophila. *Science (New York, N.Y.)*, **367**, 1230–1234.
185. Cao, Q., Boyer, D.R., Sawaya, M.R., Ge, P., Eisenberg, D.S., (2019). Cryo-EM structures of four polymorphic TDP-43 amyloid cores. *Nature Struct. Mol. Biol.*, **26**, 619–627.
186. Guenther, E.L., Ge, P., Trinh, H., Sawaya, M.R., Cascio, D., Boyer, D.R., Gonen, T., Zhou, Z.H., et al., (2018). Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nature Struct. Mol. Biol.*, **25**, 311–319.
187. Ragonis-Bachar, P., Landau, M., (2021). Functional and pathological amyloid structures in the eyes of 2020 cryo-EM. *Curr. Opin. Struct. Biol.*, **68**, 184–193.
188. Ghosh, U., Yau, W.M., Tycko, R., (2018). Coexisting order and disorder within a common 40-residue amyloid-beta fibril structure in Alzheimer's disease brain tissue. *Chem. Commun.*, **54**, 5070–5073.
189. Wiltzius, J.J.W., Landau, M., Nelson, R., Sawaya, M.R., Apostol, M.I., Goldschmidt, L., Soriaga, A.B., Cascio, D., et al., (2009). Molecular mechanisms for protein-encoded inheritance. *Nature Struct. Mol. Biol.*, **16**, 973.
190. Landau, M., Sawaya, M.R., Faull, K.F., Laganowsky, A., Jiang, L., Sievers, S.A., Liu, J., Barrio, J.R., et al., (2011). Towards a pharmacophore for amyloid. *PLoS Biol.*, **9**, e1001080
191. Lu, J.X., Qiang, W., Yau, W.M., Schwieters, C.D., Meredith, S.C., Tycko, R., (2013). Molecular structure of beta-amyloid fibrils in Alzheimer's disease brain tissue. *Cell*, **154**, 1257–1268.
192. Qiang, W., Yau, W.-M., Luo, Y., Mattson, M.P., Tycko, R., (2012). Antiparallel β-sheet architecture in Iowa-mutant β-amyloid fibrils. *PNAS*, **109**, 4443–4448.
193. Paravastu, A.K., Leapman, R.D., Yau, W.M., Tycko, R., (2008). Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *PNAS*, **105**, 18349–18354.
194. Paravastu, A.K., Petkova, A.T., Tycko, R., (2006). Polymorphic fibril formation by residues 10–40 of the Alzheimer's beta-amyloid peptide. *Biophys. J.*, **90**, 4618–4629.
195. Tycko, R., Wickner, R.B., (2013). Molecular structures of amyloid and prion fibrils: consensus versus controversy. *Acc. Chem. Res.*, **46**, 1487–1496.
196. Anfinsen, C.B., (1972). The formation and stabilization of protein structure. *Biochem. J.*, **128**, 737.
197. Anson, M.L., Mirsky, A.E., (1925). On some general properties of proteins. *J. Gen. Physiol.*, **9**, 169–179.
198. Mirsky, A.E., Pauling, L., (1936). On the structure of native, denatured, and coagulated proteins. *PNAS*, **22**, 439.
199. Pinheiro, F., Santos, J., Ventura, S., (2021). AlphaFold and the amyloid landscape. *J. Mol. Biol.*,, 167059.
200. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., et al., (2017). Mastering the game of Go without human knowledge. *Nature*, **550**, 354–359.
201. Tegunov, D., Xue, L., Dienemann, C., Cramer, P., Mahamid, J., (2021). Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nature Methods*, **18**, 186–193.
202. Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., et al., (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.*, **2**, 1–13.
203. X. Du, H. Wang, Z. Zhu, X. Zeng, Y.-W. Chang, J. Zhang, E. Xing, M. Xu, Active learning to classify macromolecular structures in situ for less supervision in cryo-electron tomography, ArXiv Preprint ArXiv:2102.12040. (2021). https://doi.org/10.1093/bioinformatics/btab123.
204. AlQuraishi, M., (2019). End-to-end differentiable learning of protein structure. *Cell Syst.*, **8**, 292–301.
205. Ingraham, J., Riesselman, A.J., Sander, C., Marks, D.S., (2018). Learning protein structure with a differentiable simulator. *International Conference on Learning Representations (ICLR)*,.