RESEARCH ARTICLE

WILEY PROTEINS

Critical assessment of methods of protein structure prediction (CASP)—Round XII

John Moult[1] | Krzysztof Fidelis[2] | Andriy Kryshtafovych[2] |

Torsten Schwede[3] | Anna Tramontano[4]

[1]Institute for Bioscience and Biotechnology Research and Department of Cell Biology and Molecular Genetics, University of Maryland, 9600 Gudelsky Drive, Rockville, Maryland 20850

[2]Genome Center, University of California, Davis, 451 Health Sciences Drive, Davis, California 95616

[3]University of Basel, Biozentrum & SIB Swiss Institute of Bioinformatics, Basel, Switzerland

[4]Department of Physics and Istituto Pasteur - Fondazione Cenci Bolognetti, Sapienza University of Rome, P.le Aldo Moro, 5, Rome 00185, Italy

Correspondence
John Moult, Institute for Bioscience and Biotechnology Research and Department of Cell Biology and Molecular Genetics, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850
Email: jmoult@umd.edu

Funding information
US National Institute of General Medical Sciences, Grant Number: R01 GM100482

Abstract

This article reports the outcome of the 12th round of Critical Assessment of Structure Prediction (CASP12), held in 2016. CASP is a community experiment to determine the state of the art in modeling protein structure from amino acid sequence. Participants are provided sequence information and in turn provide protein structure models and related information. Analysis of the submitted structures by independent assessors provides a comprehensive picture of the capabilities of current methods, and allows progress to be identified. This was again an exciting round of CASP, with significant advances in 4 areas: (i) The use of new methods for predicting three-dimensional contacts led to a two-fold improvement in contact accuracy. (ii) As a consequence, model accuracy for proteins where no template was available improved dramatically. (iii) Models based on a structural template showed overall improvement in accuracy. (iv) Methods for estimating the accuracy of a model continued to improve. CASP continued to develop new areas: (i) Assessing methods for building quaternary structure models, including an expansion of the collaboration between CASP and CAPRI. (ii) Modeling with the aid of experimental data was extended to include SAXS data, as well as again using chemical cross-linking information. (iii) A team of assessors evaluated the suitability of models for a range of applications, including mutation interpretation, analysis of ligand binding properties, and identification of interfaces. This article describes the experiment and summarizes the results. The rest of this special issue of *PROTEINS* contains papers describing CASP12 results and assessments in more detail.

KEYWORDS
CASP, community wide experiment, protein structure prediction

1 | INTRODUCTION

1.1 | Structure of the CASP experiments

CASP provides an avenue for objective testing and assessment of protein structure modeling methods. The experiments are held biannually. The integrity of the experiments is ensured by blind testing and assessment procedures: participants do not know the answers to the modeling challenges and independent assessors do not know the identity of participants.

Information about soon to be solved structures ("targets" in CASP jargon) is collected from the experimental community and passed on to the modeling community through the Prediction Center (http://predictioncenter.org). Research groups may participate via servers using fully automated methods or as experts, where a combination of computational methods and human expertise may be used. Expert groups are usually allowed up to 3 weeks to submit a model, versus 3 days for servers. Groups are limited to a maximum of 5 models per target, and are instructed that most emphasis in the assessments will be placed on the model they designate as "model 1" (intended to be the most accurate model). Predictions must be submitted to the Prediction Center in a specified machine-readable format. Accepted submissions are issued an accession number, serving as the record that a prediction had been made by a particular group on a particular target.

The models are compared with the corresponding experimental structures using a range of numerical evaluation criteria, summarized in[1] and discussed in more detail in the assessment papers in this issue.

A key aspect of the experiments is that independent assessors are asked to interpret the results. Assessors are encouraged to base their analysis on the established CASP measures and also to develop additional measures they consider appropriate.

The CASP12 prediction period was from May till August 2016. A planning meeting was held in October, at which the assessors presented their findings to each other and to the organizers. After the assessors had reported their conclusions, group identities were revealed and the most successful groups as well as those with the most promising novel methods were invited to talk at the CASP conference. The conference was held in Gaeta, Italy, in December 2016. The program of the CASP12 meeting can be found at http://predictioncenter.org/casp12/doc/CASP12_Meeting_Program.html. Many of the conference presentations as well as all results are also available on the web site.

1.2 | CASP12 statistics: precipitating groups, targets, and submissions

CASP12 maintained the high participation level of recent CASPs with 188 methods from 96 research groups in 19 countries taking part. The number of methods decreased slightly from the 207 of CASP11, primarily as a result of the elimination of the disorder prediction category and limiting the number of methods from the same research group to five.

About 34 experimental structure determination groups provided modeling targets (see the CASP12 Targets Highlights paper[2] for details of some of these). About 82 structures were selected by the CASP organizers, resulting in a total of 90 CASP12 targets (T0859 to T0948). The number of the targets is larger than the number of unique structures as 8 were released twice: targets T0929-T0934 were re-released with more accurate information on the oligomeric state; target T0946 was a re-release of T0919, with a corrected sequence; and target T0948 was a re-release of T0916 so that all groups had an opportunity to make use of a newly available homologous structure. Eleven targets were eventually canceled because the experimental structures were not available in time, leaving 71 targets for the assessment. Supporting Information Figure 1 shows the average difficulty of these targets and that of earlier CASPs. (CASP classifies the difficulty of targets based on sequence and structural similarity to the closest available structural template.[3])

CASP assesses modeling methods in a number of different categories, and these continue to evolve as the field changes. In CASP12, all 71 targets were assessed in the tertiary structure, contact prediction, and model accuracy categories. Thirty targets were evaluated in the quaternary assembly category, including 8 hetero-multimeric complexes. About 42 targets were selected for the refinement category. There were 10 targets for SAXS-assisted modeling and 3 targets for cross-linking assisted modeling. Tertiary structure assessment is divided into 2 categories—targets where one or more structure templates can be identified from sequence (template-based modeling, TBM), and those where there were no such templates (free modeling, FM). Some targets bridge these 2 categories and are referred to as TBM/FM. Properties of the targets are discussed in a separate article[4] in this PROTEINS issue.

In between CASP rounds, the CAMEO project complements the experiment by providing an automated continuous benchmarking platform for developers of server methods, using the weekly PDB prerelease information to identity targets. Several of the leading groups tested and benchmarked their new methods in preparation for CASP12. New CAMEO categories currently in implementation are continuous assessment of complexes (homo- and hetero oligomeric), residue-residue contact prediction, and ligand conformation in 3D structure modelling[5] (cameo3d.org).

Almost 55 thousand models were submitted in CASP12, of which 37,672 were three-dimensional coordinate sets. The remaining submissions were for refinement (6,227), estimation of model accuracy (7,400), residue–residue contacts (3,077), and data-assisted predictions (528).

1.3 | Management and organization

The CASP12 organizers were unchanged from CASP11 and are the authors of this article. They are responsible for all aspects of the experiment. There is an advisory board composed of senior members of the modeling community. A participants' meeting during each CASP conference allows for more direct interaction, including votes on issues of CASP policy. The Protein Structure Prediction Center is responsible for the experiment data management, including the distribution of target information, collection of predictions, generation of numerical evaluation data, developing tools for data analysis, data security, and maintenance of a web site where all data are available.

With great sorrow, we report that a key member of the organizing committee and one of the authors of this article, Anna Tramontano died shortly after the CASP12 meeting. Anna's many and extraordinary contributions to CASP are described in a tribute in this issue of PROTEINS[6] and elsewhere.[7]

1.4 | Assessment

In CASP12, there were 5 independent assessors. Matteo dal Peraro (EPFL, Lausanne, Switzerland) assessed the template free and data-assisted modeling categories; Alexandre Bonvin (University of Utrecht, Netherlands) assessed contact prediction; and Francesco Gervasio (University College London) assessed refinement. Russ Altman (Stanford, USA) together with Sean Money (U. Washington, USA) and Aleix Lafita assessed the suitability of models for deducing aspects of function. Guido Capitani (Paul Scherrer Institut, Switzerland) began the task of assessing macromolecular assemblies. Unfortunately, ill heath prevented him from continuing and Aleix Lafita and Spencer Bliven, members of his research group, generously agreed to take over the task. Sadly, Guido Capitani died shortly after the CASP12 experiment.

In the following sections, we briefly review the results and conclusions for each category of the CASP12 experiment.

2 | RESULTS

2.1 | Prediction of three-dimensional contacts

The most notable progress in CASP12 resulted from sustained improvement in methods for predicting three-dimensional contacts
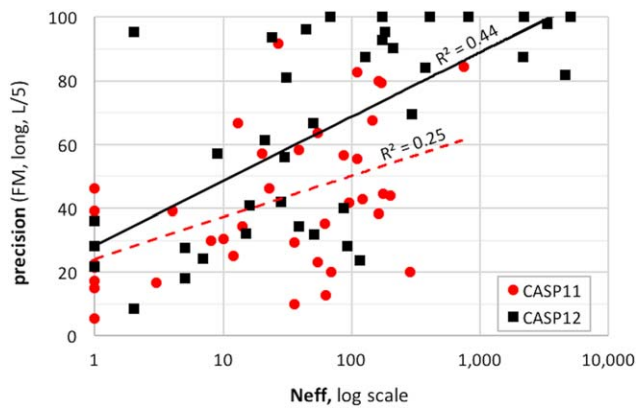
FIGURE 1 Contact prediction accuracy in CASPs 11 and 12 against effective alignment depth. As expected, accuracy increases with alignment depth, and for a number of CASP12 targets with deep alignments, precision is 100%. Best results on the set of free modeling targets are shown. Precision is for the most confidently predicted L/5 contacts separated by >23 residues in the sequence, where L is the target length. Neff is the number of diverse (<90% ID) homologous sequences covering at least 60% of the target with an E-score of 10^{-3} or better, retrieved by HHblits from the uniprot20 database

between pairs of residues in structures. In the previous CASP (11), there were just 2 targets where new approaches to this problem[8] resulted in accurate models for relatively large structures for which no template was available.[9] In CASP12, most groups had adopted the new methods for contact prediction, with impressive results: for the best performing group, precision for the most confidently predicted L/5 contacts (where L is the length of the target protein) between residues 24 or more apart in the sequence increased from 27% to 47%.[10] Figure 1 shows the relationship between L/5 precision and the depth of sequence alignment. Accuracy is 100% for some targets with deep alignments. Alignment depths are rapidly increasing because of genome sequencing projects, and so we expect that in future these methods will have ever increasing usefulness. Full details of the assessment are available in.[10] Over the history of CASP, this is the clearest and most dramatic example of a theoretical advance leading to improvements in model accuracy.

2.2 | Template free modeling

As a result of the improvements in contact prediction described above there are also dramatic improvements in accuracy for nontemplate-based models. Figure 2 shows the backbone accuracy of the best models submitted in this category for the 3 most recent CASPs. Topologically accurate models (typically above a threshold of around 50 GDT_TS units[9]) for targets shorter than 100 residues are primarily a result of the use of fragment assembly methods developed some while ago.[11] But at >100 residues, this quality of models was almost never seen until now: in CASP12, 50% of the 32 targets longer than 100 residues have better than 50 GDT_TS accuracy, with many substantially higher than that. Figure 3 shows the relationship between CASP12 best GDT_TS score for each target and the alignment depth. It

is immediately apparent that accurate models are much more likely to be produced when deep alignments are available, supporting the conclusion that improved accuracy is primarily because of more accurate contact prediction.

Figure 4 shows an example of a high accuracy model resulting from high accuracy contact prediction. This target had the deepest available sequence alignment (5095 by HHblits[12]), and the accuracy for the most confident L/5 contacts was 100%. The model has a GDT_TS of 80, compared with an average of 39 for 20 server models. The authors of this model report that a contributing factor to the high accuracy was using the predicted contacts to identify structural templates that could not be found by sequence. Other factors may also have contributed to improved CASP12 performance in this category. Another group with a successful model for this target report that they were able to identify less accurate parts of their initial contact assisted structure using model accuracy estimate methods, allowing them to concentrate on remodeling those regions. Two groups report that more accurate identification of domain boundaries made a significant difference, although incorrect domain definitions were still the biggest cause of poor models. Extensive analysis of CASP12 free modeling performance can be found elsewhere in this issue.[13]

2.3 | Template-based modeling

In spite of the impressive improvements in template free modeling outlined above, models based on templates identified by sequence similarity remain the most accurate. Over the course of the CASP experiments there have been enormous improvements in this area, as a result of
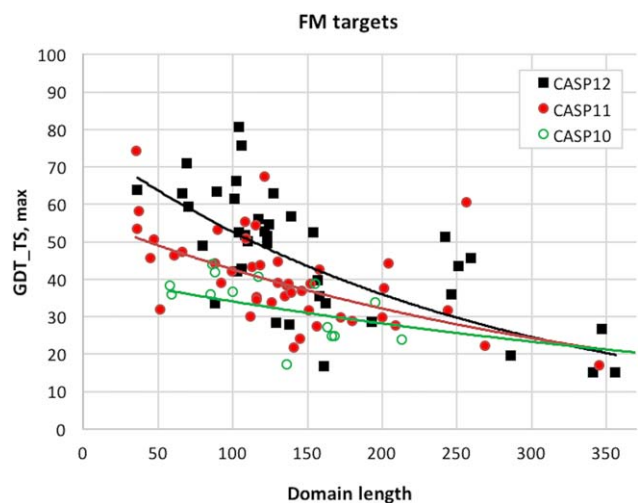
FIGURE 2 Backbone accuracy (GDT_TS) of the best submitted models in the free modeling category for the 3 most recent CASPs, as a function of target length. Good performance for targets smaller than 100 residues mostly reflects earlier improvements in this category. In CASP10, no models longer than 100 residues had GDT_TS >50. In CASP11, 4 crossed this threshold. In CASP12, half of the targets longer than 100 residues do so. (On the GDT_TS scale, 100 is perfect agreement with experiment, 20–30 is typically random, and structures with scores above 50 are largely topologically correct.)

CASP12, FM targets
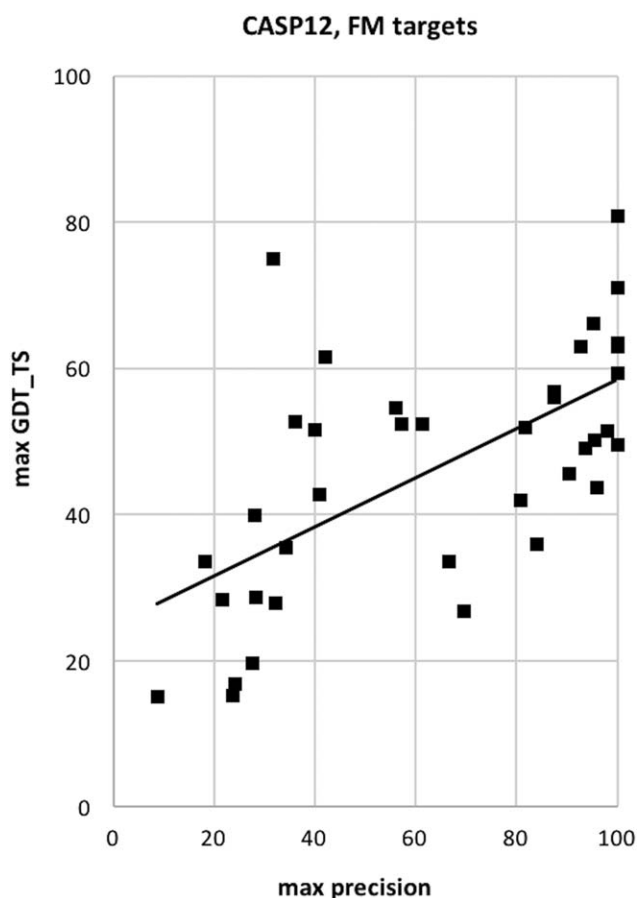
FIGURE 3 Relationship between highest backbone accuracy (GDT_TS) and highest contact prediction accuracy for free modeling targets in CASP12. Average structure accuracy doubles as contact accuracy increases, demonstrating that high accuracy is a consequence of the availability of largely correct contacts. (Precision is for the $L/5$ most confidently predicted contacts separated by at least 23 residues in the sequence, L is target length)

3 primary advances: improved alignment methods, the use of multiple structural templates, and improved nontemplate-based modeling methods. But in recent CASP experiments, overall accuracy improvement had slowed. In CASP12, however, there was a burst of progress (Figure 5).

Several factors contributed to this, including more accurate alignment of the target sequence to that of the most useful template, and improved accuracy of regions not covered by the principal template (see data in the template based modeling assessment article[15]). Considerable effort has been put into developing methods that can make use of information from multiple templates, and there is evidence that in some structures multiple templates are being successfully utilized.[15] Multiple templates also contribute to the improvement in nonprincipal template regions. There may also be improvement in nontemplate "loop" modeling methods but this is not easy to detect. Many groups now refine initial models in a variety of ways, and as discussed in the refinement section, these methods are becoming more effective. As in the FM category, some of the successful TBM modeling groups are using methods that estimate model accuracy as an aid in selecting

models and parts of models. One factor which does not appear to directly contribute to improved TBM models is contact prediction. As Supporting Information Figure 2 shows, GDT_TS as a function of alignment depth is approximately constant for template-based models. Figure 6 shows an example of an outstandingly accurate model of a difficult TBM/FM target, T0868, showing several substantial improvements over the structure that could be obtained by copying a single template. As discussed in the TBM assessment article,[15] success in this case is probably because of several factors, including selection of a nonobvious template, modeling of nontemplate modeled regions using fragment assembly methods, and refinement.

2.4 | Estimating model accuracy (EMA)

Models are still rarely as accurate as X-ray structures so that it is essential to provide users with useful estimates of accuracy, both globally and for parts of a structure. As noted above, such estimates are now also in extensive use in constructing models, assisting identification of low accuracy structures and regions in need of remodeling.

Assessment of model accuracy estimates has been included in CASP since CASP7 (2006)[16] and steady improvement in the accuracy of the methods was reported in CASPs 8, 9, 10, and 11.[17–20] CASP12 results continue that trend, with measurable progress in almost all areas of the assessment (see the EMA assessment article, this issue[21]). Of particular note are improvements in "single model methods"—those that do not require clustering analysis of a set of models. In every-day practice sets of models are not usually available, and thus single model approaches are more useful. Advances were made in picking the most
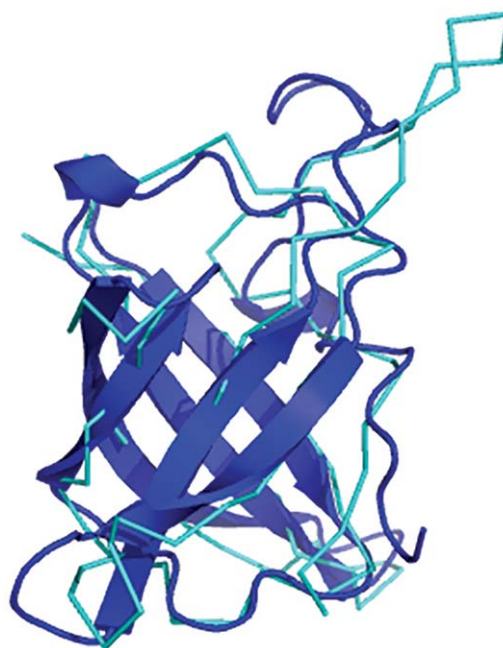
FIGURE 4 Superposition of the best model received for target T0866, the periplasmic domain of MlaD from *E.coli* (blue), with the corresponding experimental structure (turquoise, PDB 4cx8). There were no sequence detectable templates for this protein, and the outstandingly accurate model is largely because of successful prediction of a set of three-dimensional contacts

TBM and TBM/FM targets
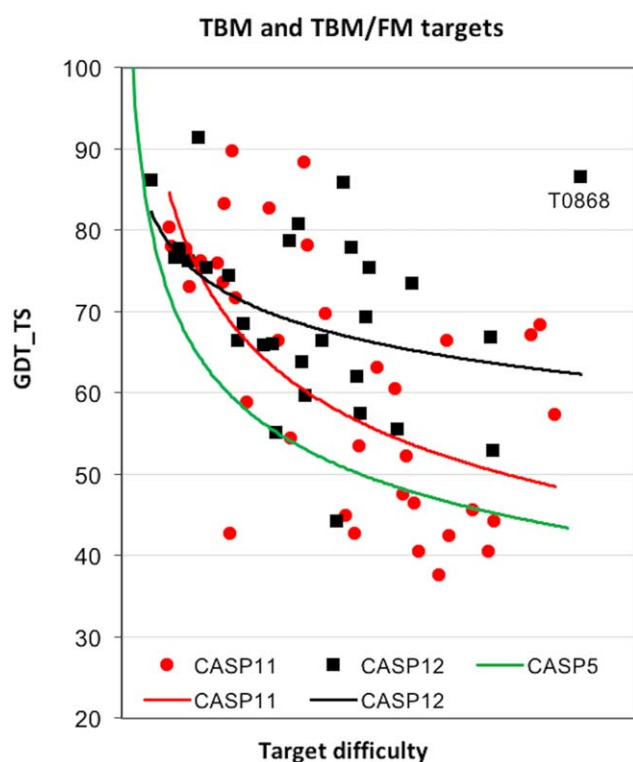
FIGURE 5 Trend lines for best model backbone accuracy (by GDT_TS) in CASP5 (2002), CASP11, and the most recent CASP12, for the template-based modeling targets (TBM and TBM/FM). By this measure, there was only modest improvement in 12 years between CASP5 and 11, but a substantial jump in the last 2 years. Points show the CASP11 and CASP12 best models for each target. The case of T0868 is discussed in the text and shown in Figure 6. The "Target Difficulty" rank of each target is based on its sequence and structure similarity to the closest template[14]

accurate models in a set, identifying unreliable regions in a model, and assigning per residue error estimates. For model selection, single model methods are now more accurate than clustering methods, with an average error of only 5 GDT_TS units between selected models and most accurate ones (Figure 7).

Target T0866 is an example of successful selection of the most accurate model by single-model EMA methods (Supporting Information Figure 3). Server models for this target span a wide range of accuracy ($13 < GDT\_TS < 73$), with most models (97%) having a score <50. In these circumstances, clustering methods do badly because they partly rely on departure from consensus to detect errors. Two single-model EMAs selected the most accurate model, and 3 more had an error of only 1.5 GDT_TS units. The main sources of single-model method advances appear to be energy function improvements and enhancement of machine learning techniques.

2.5 | Data-assisted modeling

Data-assisted or hybrid modeling, in which low-resolution experimental data are combined with computational methods, is becoming increasing important for a range of experimental data, including NMR, chemical cross-linking and surface labeling, X-ray and neutron scattering, and electron microscopy.[22] Previous CASPs have included challenges using artificial NMR NOE and other cross-link data.[12] In CASP11 there was a challenge using experimental chemical cross-linking data, provided by Adam Belsom (University of Edinburgh) and Juri Rappsilber (UofE and Berlin Technical University). In CASP12, this group provided cross-linking data for 3 targets.[23] For the first time in CASP, we were able to provide experimental SAXS data, for 12 targets.[23] This experiment was made possible by a collaboration with John Tainer's group at Lawrence Berkeley National Laboratory (Susan Tsutakawa, Greg Hura, Kathryn Burnett, and Tadeusz Ogorzalek) as well as the use of the DOE Advanced Light Source at LBNL.

In CASP11, interpretation of the modeling results with cross-linking data was complicated by false positives in the data.[24] The experimental data provided for CASP12 were much more accurate. A total of 11 groups submitted both cross-link assisted and unassisted models. Only 1 group appeared to be able to significantly enhance model accuracy using the cross-link information, but interpretation of that result is complicated by several factors (see the data assisted modeling assessment article for details[25]). As often in CASP, we will have to wait for the next experiment to see if this promising result is sustained. There are 2 possible reasons why the cross-linked data seem to be difficult to utilize. First, because of the chemistry used, backbone atoms of cross-linked residues may be up to 25 Å apart. For small proteins that may not restrict the possible conformations very much. Second, crosslinks were not evenly distributed throughout the structures.[25]
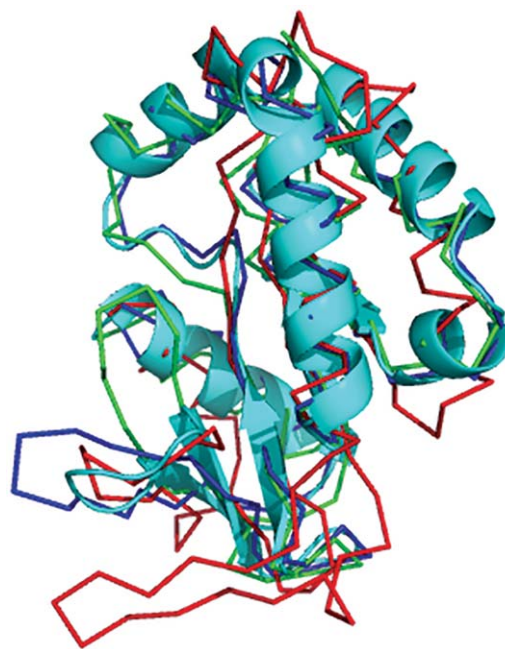
FIGURE 6 Example of accurate template-based modeling for a relatively difficult target, T0868, a bacterial CdiA tRNase toxin. The experimental structure (PDB 5j4a) is shown as a cyan cartoon, with the best homologous template in red, the best server model in green, and the best overall model in blue. There are several obvious areas of improvement over the template, for example modeling of the top left helix, not present in the template, correction of the inter-helical relationship on the top right, and correct replacement of the long template hairpin at the bottom of the structure
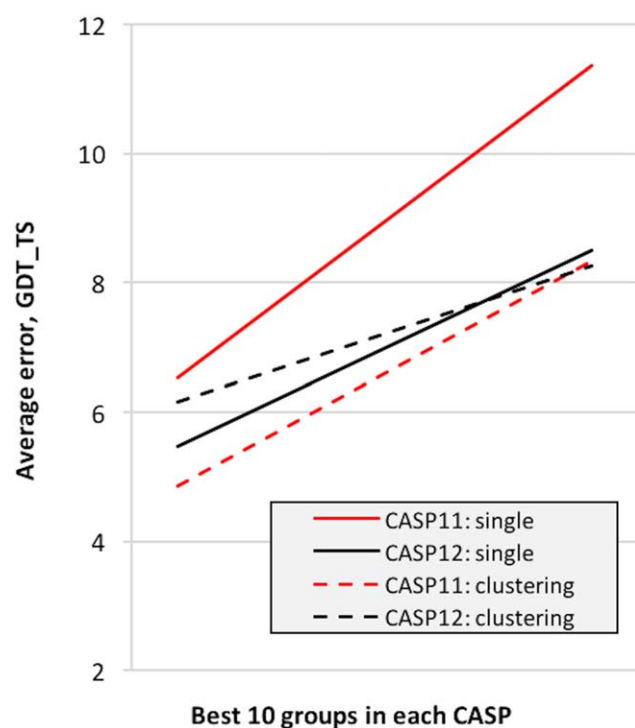
FIGURE 7 Trend lines for average error (GDT_TS units) in identifying the best model for CASP11 (red) and CASP12 (black) targets. Lower lines indicate smaller error and thus better performance. The results of the top 10 "single-model" methods (solid lines) are significantly better in CASP12 (black) than in CASP11 (red). In CASP12, the accuracy of the best single-model methods (black line) is higher than that of clustering methods (black dashed line), while in CASP11 the accuracy of single-model methods (red solid line) was much worse than the accuracy of clustering methods (red dashed line)

SAXS has the capacity to provide additional constraints to filter models or to provide constraints to help guide optimization of structure models, and so is a potentially powerful way of improving model accuracy. In practice, things can be more complicated. The CASP12 SAXS data were collected on the full-length proteins. In principle, this provides a realistic view of the structure to be modeled. However, many of the SAXS-aided targets in CASP12 contained substantial disordered regions which are not part of the structure observed by crystallography. The SAXS derived molecular envelopes were considerably larger than those from X-ray, and so were difficult to utilize effectively. Additionally, the multimeric states of many of the targets complicated the use of the SAXS data for free modeling targets. For these or other reasons, in CASP12 we did not observe improvement in model accuracy with the use of the SAXS data. In the next CASP, as discussed separately,[25] preference will be given to monomeric targets with fully ordered constructs.

2.6 | Refinement

The refinement category was originally introduced in CASP8 (2008)[26] to encourage the development of methods that can move an initial model away from a structural template toward a more accurate representation of the corresponding experimental structure. In this, after a difficult start, it has been successful: In the last 2 CASPs in particular, there have been many impressive examples of refinement of server starting models.[27,28] From a practical point of view, it is important to also have methods that perform consistently. While there has been progress there too, it is still the case that some of the more aggressive refinement methods will occasionally move substantially away from the experimental structure rather than toward it, and this is more of an issue with the CASP12 results than in the previous CASP. Related to that, as has been observed before in CASP, some targets seem more amenable to refinement than others—all the top performing methods can improve accuracy on some targets, for example T0947 and T0948, while for other targets all methods tend to decrease accuracy (for example T0879). The reasons for this are not yet clear, but a contributing factor is environmental effects—missing ligands, crystal contacts, and contacts with parts of the target not included. It might be expected that methods for estimating local and global model errors (the EMAs discussed above) would be useful in allowing participants to detect when a refinement has failed, but either the methods are not subtle enough for this application or the participants are not yet applying them. Success in refinement does not seem to be strongly related to the accuracy of the starting structure.

A motivation for developing the refinement category was a belief that the limits of informatics-based methods were being reached, and that a return to physics was essential to finally have modeling methods that rival experiment. That has turned out to be partly true: as noted by the refinement assessors for CASP12,[28] 3 of the 4 top performing methods in CASP12 use molecular dynamics (MD) with a physics-based force field as part of the refinement process. However, only 1 of these methods relies solely on MD, and the others use a wide variety of conformational sampling techniques, including fragment-based assembly, normal modes, and secondary structure remodeling. In fact, it has become difficult to distinguish the techniques used in the refinement category from those in the template based and free modeling categories. Conversely, a number of TBM and FM pipelines now incorporate MD refinement. Thus, while the refinement problem is clearly only partly solved, it is becoming harder to devise effective ways of assessing progress in the CASP context.

2.7 | Protein assemblies

The majority of proteins in living cells exhibit some form of higher-order structure in their biologically relevant state and function is often coupled to the formation of macromolecular complexes.[29,30] Not surprisingly, mutations at protein-protein interfaces with structural effects destabilizing or altering the interactions are often associated with diseases[31] and structure information has been shown to be valuable in interpreting the functional effects of mutations on protein interactions.[32] Prediction of the correct quaternary structure of a protein is therefore an essential aspect of protein structure modelling in order to provide models useful for biological applications. As part of the TBM and FM tertiary structure categories, CASP participants are expected to submit models in the correct quaternary structure state. Since

CASP9, assessors have included evaluation of homomeric assemblies, but with limited success because of the small number of CASP groups systematically predicting assemblies, and the overall low quality of the predictions.[33] In round 11, CASP increased the emphasis on assembly structures by extending the targets to include heteromeric complexes and by collaborating with CAPRI to address this area.[34,35] The area was further strengthened in CASP12 by introducing a separate category in which the accuracy of the predicted complexes for all nonmonomeric targets was evaluated by a dedicated team of assessors,[36] and a subset of targets meeting CAPRI criteria was separately assessed in collaboration with CAPRI.[37]

A total of 30 oligomeric targets were suitable for assembly prediction and assessment: 8 heteromeric assemblies and 22 homo-oligomeric targets. There was a significant increase in the number of groups submitting models of complexes, with 68 groups submitted models for at least 3 distinct oligomeric targets. However, only 10 groups submitted models for more than 10 targets. As expected, participation and accuracy was higher for target assemblies suitable for homology modelling, while prediction of protein interface contacts without a template proved more challenging. Interestingly, for most targets the best predictions outperform a sequence-based naïve homology modelling approach.[36]

Although there is overall progress in the prediction of protein assemblies in CASP12 compared to previous experiments, there is still much room for improvement. For instance, only a few groups (mostly servers) systematically modeled complexes, and co-evolution information for predicting interface contacts seems to have not yet led to the same success as in tertiary structure prediction. We hope that continued emphasis in this area will encourage methods developers to address this biologically important aspect of structure modeling.

2.8 | Function analysis

As noted above, in spite of much progress, computational models are still seldom as accurate as X-ray experimental structures. But how accurate is accurate enough? Experimental structures are almost always solved with the expectation that the results will contribute to understanding some aspect of function or will aid in developing some application. Different objectives require different levels of accuracy[38,39]: at one extreme, identifying peptides that might be used as epitopes in eliciting an immune response requires only an approximate structure, while at the other, drug design will probably require multiple high-resolution ones. CASP has launched a new category of assessment aimed at encouraging the development of methods that can determine whether a model is adequate for answering a particular biological question. In CASP11, Roland Dunbrack (Fox Chase Cancer Center, USA) pioneered this form of analysis.[40] Building on those results, in CASP12, Russ Altman (Stanford, USA) led a collaborative effort in this area,[41] examining regions suggested by the target providers as most functionally relevant and other areas with known functional involvement. His group applied methods of analyzing the structural environment of a site[42] to determine whether ligand binding properties or the effect of a mutation can be reliably deduced from a particular model. Sean

Mooney (U. Washington, USA) used analysis of structural features to see if human genetic variants in CASP models could be reliably interrupted and Aleix Lafita (Paul Scherrer Institut, Switzerland) examined the accuracy of pockets that span subunit interfaces. In general, as one might expect, there is a correlation between model accuracy and accuracy of these regions, though there are exceptions, suggesting that calibrating this type of approach will be worthwhile. As is often the case in CASP, it will probably take several more rounds for the category to mature.

3 | DISCUSSION

As outlined above, this round of CASP saw substantial progress in 4 areas—contact prediction, free modeling, template-based modeling, and estimating the accuracy of models. That follows the long-term trend in CASP so that, cumulatively, the accuracy of modeling methods has increased enormously (see Supporting Information Figure 4). Some of this improvement is because of increased availability of data, both sequence and structure. The modeling difficulty scale used to compare results within and across CASPs partly normalized for that, but only partly, since it does not take into account how many related sequences and structures are available for a target, only the usefulness of the closest ones. Much of the progress is from the development of methods that use multiple sequence information to improve alignments, for example,[43] and contact prediction and methods that use multiple structure information to choose fragments and substructures for particular regions. The history of contact prediction[8] demonstrates that developing effective algorithms for making use of this information is often demanding.

As the field has matured it has also become of great practical use—servers such as Swissmodel[44] handle approximately 1 request for a three-dimensional model every minute. This is nice to see, but it also changes the landscape by which methods should be judged—a method may become more accurate overall, but if that accuracy varies from model to model and cannot be estimated, the advance is of only academic interest. We see both sides of this issue in CASP12. On the one hand, although the power of refinement methods continues to increase, greater inconsistency in performance in CASP12 means that it is difficult to yet recommend their use in routine modeling. On the other hand, the increasing utility of "single model" methods for estimating model accuracy implies that it should now be possible for all models to be accompanied by meaningful estimates of global and local accuracy. In turn, that should help users judge suitability for their applications.

While it is useful to consider each category of CASP separately as we have done here, that viewpoint tends to miss the power of combining methods. As we have seen, improved contact methods lead to improved FM model accuracy, fragment assembly methods originally developed for FM modeling are apparently contributing to nontemplate modeling, molecular dynamics refinement is integrated into FM and TBM pipelines, and it appears that accuracy estimate methods are helping judge which parts to remodel. In this respect, progress in CASP

continues to arise from both new algorithms and the engineering needed to maximally exploit those advances.

The categories included in CASP continue to change in response to the evolution of the field and also to encourage new directions. In this round we emphasized 2 new areas introduced in CASP11: modeling of protein assemblies and evaluating the suitability of models for interpreting aspects of function. The protein assembly category was again conducted in collaboration with CAPRI, this time with assessment by both CASP and CAPRI, using different metrics. As a result, there are 2 assessment papers in the PROTEINS issue.

We are planning to hold CASP13 in 2018 on the same timetable as previous rounds, with a prediction season in the late Spring, and culminating in a meeting at the end of the year. Details will be posted on the Prediction Center web site as they become available. Those interested may also register on the web site, and so receive updates.

ORCID

John Moult http://orcid.org/0000-0002-3012-2282
Krzysztof Fidelis http://orcid.org/0000-0002-8061-412X
Andriy Kryshtafovych http://orcid.org/0000-0001-5066-7178
Torsten Schwede http://orcid.org/0000-0003-2715-335X

REFERENCES

[1] Kryshtafovych A, Monastyrskyy B, Fidelis K. CASP11 statistics and the prediction center evaluation system. Proteins. 2016;84(suppl 1): 15–19.

[2] Kryshtafovych, Albrecht R, Baslé A, et al. Target highlights from the first post-PSICASP experiment (CASP12, May-August 2016). Proteins. 2017; CASP12 issue.

[3] Kryshtafovych A, Fidelis K, Moult J. CASP10 results compared to those of previous CASP experiments. Proteins. 2014;82:164–174.

[4] Abriata LA, Kinch LN, Tamo GE, Monastyrskyy B, Kryshtafovych A, Dal Peraro M. Definition and classification of evaluation units for tertiary structure prediction in CASP12 facilitated through semi-automated metrics. Proteins. 2018;86:000–000. https://doi.org/10.1002/prot.25403.

[5] Haas J, Barbato A, Studer G, Behringer D, Roth S, Mostaguir K, Bertoni M, Schwede T. Continuous automated model evaluation (CAMEO) complementing the critical assessment of techniques for structure prediction. Proteins. 2017;CASP12 issue.

[6] Moult J, Fidelis K, Kryshtafovych A, Schwede T. A tribute to Anna Tramontano. Proteins. 2017;CASP12 issue.

[7] Thornton JM, Valencia A, Schwede T. Anna Tramontano 1957–2017. Nat Struct Mol Biol. 2017;24(5):431–432.

[8] de Juan D, Pazos F, Valencia A. Emerging methods in protein co-evolution. Nat Rev Genet. 2013;14(4):249–261.

[9] Kinch LN, Li W, Monastyrskyy B, Kryshtafovych A, Grishin NV. Evaluation of free modeling targets in CASP11 and ROLL. Proteins. 2016;84:51–66.

[10] Schaarschmidt J, Monastyrskyy B, Kryshtafovych A, Bonvin AM. Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age. Proteins. 2017;CASP issue.

[11] Bonneau R, Baker D. Ab initio protein structure prediction: Progress and prospects. Annu Rev Biophys Biomol Struct. 2001;30(1):173–189.

[12] Remmert M, Biegert A, Hauser A, Soding J. HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods. 2011;9(2):173–175.

[13] Abriata LA, Tamo GE, Monastyrskyy B, Kryshtafovych A, Dal Peraro M. Assessment of hard target modeling in CASP12 reveals an emerging role of alignment-based contact prediction methods. Proteins. 2018;86:000–000. https://doi.org/10.1002/prot.25423

[14] Kryshtafovych A, Fidelis K, Moult J. CASP9 results compared to those of previous CASP experiments. Proteins. 2011;79(S10):196–207.

[15] Kryshtafovych A, Monastyrskyy B, Fidelis K, Moult J, Tramontano A. Evaluation of template-based modeling in CASP12. Proteins. 2017;CASP12 issue.

[16] Cozzetto D, Kryshtafovych A, Ceriani M, Tramontano A. Assessment of predictions in the model quality assessment category. Proteins. 2007;69(S8):175–183.

[17] Cozzetto D, Kryshtafovych A, Tramontano A. Evaluation of CASP8 model quality predictions. Proteins. 2009;77(S9):157–166.

[18] Kryshtafovych A, Fidelis K, Tramontano A. Evaluation of model quality predictions in CASP9. Proteins. 2011;79(suppl10):91–106.

[19] Kryshtafovych A, Barbato A, Fidelis K, Monastyrskyy B, Schwede T, Tramontano A. Assessment of the assessment: Evaluation of the model quality estimates in CASP10. Proteins. 2014;82(Suppl2):112–126.

[20] Kryshtafovych A, Barbato A, Monastyrskyy B, Fidelis K, Schwede T, Tramontano A. Methods of model accuracy estimation can help selecting the best models from decoy sets: Assessment of model accuracy estimations in CASP11. Proteins. 2016;84:349–369.

[21] Kryshtafovych A, Monastyrskyy B, Fidelis K, Schwede T, Tramontano A. Assessment of model accuracy estimations in CASP12. Proteins. 2017;CASP12 issue.

[22] Sali A, Berman HM, Schwede T, et al. Outcome of the first wwPDB hybrid/integrative methods task force workshop. Structure. 2015;23 (7):1156–1167.

[23] Ogorzalek TL, Hura GL, Belsom A, Burnett KH, Kryshtafovych A, Tainer JA, Rappsilber J, Tsutakawa SE, Fidelis K. Small angle X-ray scattering and cross-linking for data assisted protein structure prediction in CASP 12 with prospects for improved accuracy. *Proteins.* 2017;CASP12 issue.

[24] Schneider M, Belsom A, Rappsilber J, Brock O. Blind testing of cross-linking/mass spectrometry hybrid methods in CASP11. *Proteins.* 2016;84:152–163.

[25] Dal Peraro M. al. e. [data assisted assessment paper. PROTEINS 2017;CASP12 issue.

[26] MacCallum JL, Hua L, Schnieders MJ, Pande VS, Jacobson MP, Dill KA. Assessment of the protein-structure refinement category in CASP8. *Proteins.* 2009;77(S9):66–80.

[27] Modi V, Dunbrack RL Jr. Assessment of refinement of template-based models in CASP11. *Proteins.* 2016;84(Suppl 1):260–281.

[28] Hovan L, Oleinikovas V, Yalinca H, Kryshtafovych A, Saladino G, Luigi Gervasio F. Assessment of the model refinement category in CASP12. *Proteins.* 2018;86:000–000. https://doi.org/10.1002/prot. 25409.

[29] Krogan NJ, Cagney G, Yu H, et al. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature.* 2006;440 (7084):637–643.

[30] Orchard S, Ammari M, Aranda B, et al. The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 2014;42(Database issue):D358–D363.

[31] David A, Razali R, Wass MN, Sternberg MJ. Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Hum Mutat.* 2012;33(2):359–363.

[32] Brender JR, Zhang Y, Jernigan RL. Predicting the effect of mutations on protein-protein binding interactions through structure-based interface profiles. *PLoS Comput Biol.* 2015;11(10):e1004494

[33] Mariani V, Kiefer F, Schmidt T, Haas J, Schwede T. Assessment of template based protein structure predictions in CASP9. *Proteins.* 2011;79(Suppl 10):37–58.

[34] Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins.* 2016;84:4–14.

[35] Lensink MF, Velankar S, Kryshtafovych A, et al. Prediction of homo-protein and heteroprotein complexes by protein docking and template-based modeling: A CASP-CAPRI experiment. *Proteins.* 2016;84(Suppl 1):323–348.

[36] Lafita A, Bliven S, Kryshtafovych A, et al. Assessment of protein assembly prediction in CASP12. *Proteins.* 2018;86:000–000. https://doi.org/10.1002/prot.25408.

[37] Lensink MF, Velankar S, Baek M, Heo L, Seok C, Wodak SJ. The challenge of modeling protein assemblies: The CASP12-CAPRI experiment. *Proteins.* 2017;CASP12 issue.

[38] Baker D, Sali A. Protein structure prediction and structural genomics. *Science.* 2001;294(5540):93–96.

[39] Moult J. Comparative modeling in structural genomics. *Structure.* 2008;16(1):14–16.

[40] Huwe PJ, Xu Q, Shapovalov MV, Modi V, Andrake MD, Dunbrack RL Jr. Biological function derived from predicted structures in CASP11. *Proteins.* 2016;84:370–391.

[41] Liu T, Ish-Shalom S, Wen T, Lafita A, Bock C, Mort M, Cooper DN, Bliven S, Capitani G, Mooney SD, Altman RB. Biological and functional relevance of CASP predictions. *Proteins.* 2017;CASP12 issue.

[42] Liu T, Altman RB, Beard DA. Using multiple microenvironments to find similar ligand-binding sites: Application to kinase inhibitor binding. *PLoS Comput Biol.* 2011;7(12):e1002326

[43] Hildebrand A, Remmert M, Biegert A, Soding J. Fast and accurate automatic structure prediction with HHpred. *Proteins.* 2009;77(S9): 128–132.

[44] Bienert S, Waterhouse A, de Beer TA, et al. The SWISS-MODEL repository-new features and functionality. *Nucleic Acids Res.* 2017; 45(D1):D313–D319.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.