PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

CASP Introduction—Round IX

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

John Moult,[1*] Krzysztof Fidelis,[2] Andriy Kryshtafovych,[2] and Anna Tramontano[3]

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

[2] Genome Center, University of California, Davis, California 95616

[3] Department of Physics and Istituto Pasteur-Fondazione Cenci Bolognetti, Sapienza University of Rome, Rome 00185, Italy

ABSTRACT

This article is an introduction to the special issue of the journal PROTEINS, dedicated to the ninth Critical Assessment of Structure Prediction (CASP) experiment to assess the state of the art in protein structure modeling. The article describes the conduct of the experiment, the categories of prediction included, and outlines the evaluation and assessment procedures. Methods for modeling protein structure continue to advance, although at a more modest pace than in the early CASP experiments. CASP developments of note are indications of improvement in model accuracy for some classes of target, an improved ability to choose the most accurate of a set of generated models, and evidence of improvement in accuracy for short "new fold" models. In addition, a new analysis of regions of models not derivable from the most obvious template structure has revealed better performance than expected.

Proteins 2011; 79(Suppl 10):1–5.

Key words: protein structure prediction; community wide experiment; CASP.

INTRODUCTION

The fraction of known protein sequences for which a corresponding experimental three-dimensional structure is available is shrinking rapidly and is now less than 1 in a 1000. For other proteins, structural information can be provided by the use of computational modeling. For well over half of known sequences, there is a detectable relationship to the sequence of one or more proteins with experimental structure, and in these cases, a model can be produced based on those relationships. Model accuracy depends primarily on how strong that sequence relationship is and how much of the structure is similar to the primary template structure. For proteins where a relationship to a known structure cannot be detected, nontemplate-based modeling, sometimes referred to as *ab initio* modeling, must be used. Over the course of the Critical Assessment of Structure Prediction (CASP) experiments, we have seen enormous progress in the accuracy of both template-based modeling (TBM) and template-free modeling (FM) (see the article comparing performance in different CASPs in this issue[1]). However, except for very high levels of sequence identity between a target structure and a related protein, modeling methods are not yet competitively accurate with experimental approaches. CASP is an organization that provides a mechanism for assessing how well the modeling methods perform, and for highlighting method improvements, so that these may be rapidly adopted. It also provides a mechanism for identifying where there are bottlenecks to progress, allowing developmental effort to be better focused where most needed. The core principle of CASP is that methods are assessed on the basis of the analysis of a

The authors state no conflict of interest.
Grant sponsor: National Library of Medicine; Grant number: LM07085; Grant sponsor: NIH Institute of General Medical Sciences; Grant number: GM072354; Grant sponsor: KAUST Award; Grant number: KUK-I1-012-43.
*Correspondence to: John Moult, Institute for Bioscience and Biotechnology Research, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850. E-mail: jmoult@umd.edu
Received 7 September 2011; Accepted 12 September 2011
Published online 19 September 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.23200

large number of blind predictions of protein structure, made before experimental data are available.

This issue of PROTEINS is devoted to articles reporting the outcome of the ninth CASP experiment (CASP9) and related activities. There have been eight previous CASP experiments, at two-year intervals from 1994 through 2008, and these were reported in previous special issues of PROTEINS.[2–9] This article outlines the structure and conduct of the experiment. It is followed by an article highlighting the most interesting CASP9 targets from the perspective of the members of the experimental community submitting targets to CASP.[10] Next is the article[11] describing the CASP9 target proteins, guidelines for splitting them into domains (evaluation units), and general principles for assigning the relative difficulty of constructing an accurate model in each case. The specific challenges in constructing the best possible three-dimensional model of a particular protein depend on a number of factors. To reflect these considerations, CASP modeling targets are divided into categories. The categories have evolved over the course of the experiments, as the capabilities of the methods have changed. This time, as in CASP8, targets were divided into two primary categories—TBM, where a relationship to one or more experimentally determined structures could be identified, providing at least one modeling template and often more, and template-free modeling, where there are either no usefully related structures or the relationship is so distant that it cannot be detected (FM).

There are articles[12,13] from the assessment teams in the TBM and FM categories. Emphasis was again placed on a third category of three-dimensional modeling—refinement of initial models. To this end, selected best models submitted from the first two categories were provided as starting structures, and participants were invited to see if they could improve these. Because of the importance of this task, separate assessors were assigned, and there is an article reporting their analysis.[14]

In addition to evaluating the overall accuracy of three-dimensional structure models, CASP also examines other key aspects of structure modeling. There are four articles assessing these in the following areas: prediction of the accuracy of a model,[15] critical to determining whether it is suitable for a particular purpose; prediction of the presence of structural disorder,[16] important since some parts of proteins do not exhibit a single three-dimensional structure under all circumstances; intramolecular contact identification,[17] a source of auxiliary information for template-free modeling; and the identification of ligand binding sites,[18] a central application of models. All these categories were also included in previous CASPs. A new feature in this experiment is analysis of accuracy of quaternary structure, discussed in the TBM assessment article.[12]

The assessment articles are followed by five articles from some of the more successful teams that developed new methods. In a departure from past practice, this time we asked contributing modeling groups to concentrate on details of the methods rather than describing the results. Finally, there is an article considering the results of this CASP experiment in the context of the previous ones and highlighting the areas where progress has been made.[1] As always, the assessors' articles are probably the most important in the issue and describe the state of the art as they found it in CASP9.

THE CASP9 EXPERIMENT

The structure of the experiment was very similar to that of the earlier ones, with three main steps:

1. Participants were required to register for the experiment in one or both of two categories: as human teams, where a combination of computational methods and investigator expertise may be used, and as servers, where methods are only computational and fully automated, so that a target sequence is sent directly to a machine. In some cases, investigators registered in both categories. In practice, the main distinction between the "human" and "server" categories now is the longer time period available between the release of a target and the model submission deadline—there are very few groups where significant human expertise is brought to bear. The longer period is used in two ways—to make use of initial models produced by the rapid server stage and to perform longer calculations.
2. Information about "soon to be solved" structures was collected from the experimental community and passed on to the modeling community. Continuing the trend of recent CASPs, nearly all targets were obtained from the structural genomics community, particularly the NIH Protein Structure Initiative Centers (the PSI, http://www.nigms.nih.gov/Research/FeaturedPrograms/PSI). The high throughput of structural genomics together with a well-integrated pipeline for target identification facilitated by the PDB again allowed us to collect in excess of 100 diverse and interesting targets.
3. Models were collected in accordance with predefined deadlines (72 h after target release for servers, typically 3 weeks after release for human groups). As in previous CASPs, groups were limited to a maximum of five models per target and were instructed that most emphasis in assessment would be placed on the model they designated as the most accurate (referred to as "model 1"). This self-ranking of model quality was also used as part of the evaluation of the state of the art in assigning relative accuracy to models. The models were compared with experiment, using numerical evaluation

techniques and expert assessment, and a meeting was held to discuss the significance of the results.

MANAGEMENT AND ORGANIZATION

CASP now has a well-evolved set of systems and procedures to facilitate data management and security and to ensure that the modeling community is informed and consulted. The principal components are

A. Organizers. The authors of this article, responsible for all aspects of the organization of the experiment and meeting.
B. The FORCASP web site (www.FORCASP.org). FORCASP provides a forum where members of the modeling community may discuss aspects of the CASP experiment and provide suggestions to the organizers.
C. Participants' meeting. During each CASP conference, there is a meeting for those taking part in the experiment, with votes on issues of CASP policy, particularly for major changes and extensions of the CASP process.
D. Independent assessors. The independent assessors have primary responsibility for judging the quality of the models received and commenting on the current state of the art.
E. The Protein Structure Prediction Center. The center is responsible for all data management aspects of the experiment, 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.
F. A recently reintroduced advisory board composed of senior members of the modeling community, who advise the organizers on all aspects of the CASP experiments and related activities.

TARGETS AND PARTICIPATION

One hundred and twenty nine protein sequences were released as modeling targets. Of these, 12 were canceled, resulting in 117 where the experimental structures were available for evaluation and assessment. These were divided into domains, each of which was treated as a separate evaluation unit. In all, 148 evaluation units were included. One target, T0549, was assessed in the disorder prediction category only.

The full target set is too large for substantial amounts of human or computer time to be devoted to each in the short period of the experiment. To address this issue, as in CASP8, targets were divided into two sets. All targets were sent to registered servers, and a subset of 60 was designated for human team prediction.

The level of participation in the CASP experiment continues to be high, with 248 groups, similar to that of CASP7 and CASP8 (233 and 253 groups, respectively) and representing a large fraction of the relevant community.

COLLECTION AND VALIDATION OF MODELS

There were a total of 86,891 models deposited in CASP9, of which 61,665 were three-dimensional co-ordinate sets. A further 1220 are sequence alignments, which were converted into co-ordinates for assessment. The remaining submissions are for residue–residue contacts (4162), structural disorder (5210), binding site identification (5666), estimation of three-dimensional model quality (7116), and refinement of initial models (1709). All predictions were submitted to the Prediction Center in a machine-readable format. Accepted submissions were issued an accession number, serving as the record that a prediction had been made by a particular group on a particular target.

NUMERICAL EVALUATION OF PREDICTIONS

Over the course of the CASP experiments, a set of numerical evaluation techniques have been developed, and these are now accepted as standards. They are unaltered since CASP8 and have been described extensively in previous articles.[9,19–22] Extensive analysis using these standard metrics is provided by the Prediction Center, first to the assessors, to assist them in their work, and then to all, via the Center web site.

In addition, the assessors are encouraged to develop their own measures to complement the established CASP ones, and in this manner, a range of new and useful metrics have been introduced. It is also desirable to introduce new evaluation measures in each experiment to reduce any tendency for optimization of methods to perform well against the inevitably somewhat artificial standards imposed by the metrics. In CASP9, a new measure developed by the FM assessment team proved partially effective in solving the long-standing problem of ranking the relative quality of low accuracy models in that category.

The numerical evaluation measures, although critical, are not generally sufficient to draw final conclusions about the quality and usefulness of modeling methods. A key principle of CASP is that primary responsibility for assessing the significance of the results is placed in the hands of independent assessors. This continues to be a major source of insight and innovation in CASP, as well as ensuring that organizer biases are not imposed on the outcome. In CASP9, the TBM assessor was Torsten

Schwede (University of Basel, Basel, Switzerland), that for template-free modeling was Nick Grishin (University of Texas, Dallas, TX), and for refinement was Ken Dill (Stony Brook University, Stony Brook, NY, USA).

MEETINGS, WEB SITE, AND PUBLICATIONS

Following the established CASP procedure, a planning meeting was held approximately 2 months after the close of the modeling season, at which the assessors presented their results to each other and to the organizers. As always, the identities of participating groups were hidden from the assessors until after those presentations, to avoid ranking bias.

The meeting to discuss the outcome of the experiment was held at the Asilomar Conference Center, California, in December 2010. In addition to sessions devoted to the outcome of the experiment in each of the modeling categories, the meeting format was adjusted to address two primary goals. First, placing greater emphasis on emerging methods, facilitated by sessions in which successful modelers described their approaches in some depth, and by a "highlights" session in which authors of recently published interesting methods papers presented their work. Second, to make maximum use of the opportunity provided by most of the structure modeling community being assembled in the same place, there was a series of informal "Solutions" sessions in which participants discussed ideas for new approaches that have the potential to significantly improve model quality. There was also a session on the increasingly important topic of RNA structure modeling. The full program can be found on the Prediction Center web site.

This issue of PROTEINS is the official report of the CASP9 experiment and the outcome of the meeting. All the modeling and assessment papers in this issue have been peer reviewed. The CASP web site (http://predictioncenter.org) provides extensive details of the targets, the predictions, and the numerical analyses.

PROGRESS IN CASP9

Early CASP experiments saw dramatic improvements from round to round. Recently, progress has been more gradual, but nevertheless, steady and cumulatively very significant. This time,[1] in TBM, there is evidence of improved accuracy for targets in the midrange of difficulty, likely attributable to improved methods of combining information from multiple templates, an area where a number of innovative developments were reported. There is also evidence for improved accuracy for models of short (less than about 100 residues) template-free modeling targets, and a greater variety of methods are now effective in this area. A major remaining challenge in TBM is producing accurate structure for those regions not easily derived from an obvious template. New analysis of performance in this area has shown that methods are more effective than previously apparent, although this remains a bottleneck area. Methods for identifying the best model from a set generated have also improved.

FUTURE DEVELOPMENTS

Several new initiatives are planned within the CASP framework. A difficulty in identifying successful emerging methods is that producing a complete protein structure model is a multistep process, so that improvements in one step may not be easily spotted. To address this, for some targets, modeling will be divided into stages, first calling for identification of templates, then releasing template information and calling for alignments to those templates, and then releasing correct alignments and calling for refinement and extra-template modeling. In this way, it should be possible to determine more precisely which methods are most effective at each step and when new methods are successful. It is hoped that this procedure will also encourage individual groups to focus on a particular problem, knowing that success will be recognized.

An ongoing issue is that some classes of modeling target are quite rare, and so too few are obtained within the short three-month modeling season of a CASP experiment. The primary classes affected are template-free modeling targets, membrane proteins, and the new area of RNA structure. To address this, for these targets, there will be a rolling CASP procedure, in which targets are released to the modeling community continuously, as they become available. In this way, it should be possible to accumulate a sufficient quantity of models over the period between experiments so as to be able to obtain a clearer measure of how well methods are performing in these areas.

There will also be a continuous process for evaluating the performance of servers, made possible through collaboration with the Protein Modeling Portal's (http://www.proteinmodelportal.org/) CAMEO system. CAMEO will collect models from servers for each weekly release of experimental structures from the PDB. CAMEO organizers have agreed to make the models they collect available for analysis using the CASP pipeline.

A CASP10 experiment is planned, beginning in the spring of 2012 and culminating in a meeting in December of that year. The meeting is planned to take place in Italy. Those interested should check the CASP web site for further announcements.

ACKNOWLEDGMENTS

The authors are grateful to the members of the experimental community, particularly the structural genomics

centers, who agreed to provide targets. Taking part required courage and commitment on the part of all the modeling groups. Once again the assessment teams worked extremely hard and effectively to extract major insights from the results. The authors again thank PROTEINS for providing a mechanism for peer-reviewed publication of the outcome of the experiment. They are also grateful to Wiley and PROTEINS for agreeing to make these special issues open access, so that all scientists may easily make use of the results. They thank Helen Berman and the PDB staff for their key role in structural genomics target processing. They also thank Professor Torsten Schwede for agreeing to join the CASP organizational team, beginning with CASP10, and look forward to working with him.

REFERENCES

1. Kryshtafovych A, Fidelis K, Moult J. CASP9 results compared to those of previous CASP experiments. Proteins 2011;79(Suppl 10):196–207.
2. Moult J, Pedersen JT, Judson R, Fidelis K. A large-scale experiment to assess protein structure prediction methods. Proteins 1995;23:ii–v.
3. Moult J, Hubbard T, Bryant SH, Fidelis K, Pedersen JT. Critical assessment of methods of protein structure prediction (CASP): round II. Proteins 1997;29(Suppl 1):2–6.
4. Moult J, Hubbard T, Fidelis K, Pedersen JT. Critical assessment of methods of protein structure prediction (CASP): round III. Proteins 1999;37(Suppl 3):2–6.
5. Moult J, Fidelis K, Zemla A, Hubbard T. Critical assessment of methods of protein structure prediction (CASP): round IV. Proteins 2001;45(Suppl 5):2–7.
6. Moult J, Fidelis K, Zemla A, Hubbard T. Critical assessment of methods of protein structure prediction (CASP)-round V. Proteins 2003;53(Suppl 6):334–339.
7. Moult J, Fidelis K, Rost B, Hubbard T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)—round 6. Proteins 2005;61(Suppl 7):3–7.
8. Moult J, Fidelis K, Kryshtafovych A, Rost B, Hubbard T, Tramontano A. Critical assessment of methods of protein structure prediction-Round VII. Proteins 2007;69(Suppl 8):3–9.
9. Moult J, Fidelis K, Kryshtafovych A, Rost B, Tramontano A. Critical assessment of methods of protein structure prediction—Round VIII. Proteins 2009;77(Suppl 9):1–4.
10. Kryshtafovych A, Moult J, Bartual SG, Fernando Bazan J, Berman H, Casteel DE, Christodoulou E, Everett JK, Hausmann J, Heidebrecht T, Hills T, Hui R, Hunt JF, Jayaraman S, Joachimiak A, Kennedy MA, Kim C, Lingel A, Michalska K, Montelione GT, Otero JM, Perrakis A, Pizarro JC, van Raaij MJ, Ramelot TA, Rousseau F, Tong L, Wernimont AK, Young J, Schwede T. Target highlights in CASP9: experimental target structures for the critical assessment of techniques for protein structure prediction. Proteins 2011;79(Suppl 10):6–20.
11. Kinch L, Shi S, Cheng H, Cong Q, Pei J, Schwede T, Grishin N. CASP9 target classification. Proteins 2011;79(Suppl 10):21–36.
12. Mariani V, Kiefer F, Haas J, Schmidt T, Schwede T. Assessment of template based predictions in CASP9. Proteins 2011;79(Suppl 10):37–58.
13. Kinch L, Shi SY, Cong Q, Cheng H, Liao Y, Grishin NV: CASP9 assessment of free modeling target predictions. Proteins 2011;79(Suppl 10):59–73.
14. MacCallum JL, Pérez A, Schnieders MJ, Hua L, Jacobson MP, Dill KA: Assessment of protein structure refinement in CASP9. Proteins 2011;79(Suppl 10):74–90.
15. Kryshtafovych A, Fidelis K, Tramontano A. Evaluation of model quality predictions in CASP9. Proteins 2011;79(Suppl 10):91–106.
16. Monastyrskyy B, Fidelis K, Moult J, Tramontano A, Kryshtafovych A. Evaluation of disorder predictions in CASP9. Proteins 2011;79(Suppl 10):107–118.
17. Monastyrskyy B, Fidelis K, Tramontano A, Kryshtafovych A. Evaluation of residue-residue contact predictions in CASP9. Proteins 2011;79(Suppl 10):119–125.
18. Schmidt T, Haas J, Cassarino TG, Schwede T: Assessment of ligand binding residue predictions in CASP9. Proteins 2011;79(Suppl 10):126–136.
19. Cozzetto D, Kryshtafovych A, Fidelis K, Moult J, Rost B, Tramontano A. Evaluation of template-based models in CASP8 with standard measures. Proteins 2009;77(Suppl 9):18–28.
20. Kryshtafovych A, Krysko O, Daniluk P, Dmytriv Z, Fidelis K. Protein structure prediction center in CASP8. Proteins 2009;77(Suppl 9):5–9.
21. Kryshtafovych A, Prlic A, Dmytriv Z, Daniluk P, Milostan M, Eyrich V, Hubbard T, Fidelis K. New tools and expanded data analysis capabilities at the protein structure prediction center. Proteins 2007;69(Suppl 8):19–26.
22. Kryshtafovych A, Milostan M, Szajkowski L, Daniluk P, Fidelis K. CASP6 data processing and automatic evaluation at the protein structure prediction center. Proteins 2005;61(Suppl 7):19–23.