Progress in protein structure prediction

Alexey G. Murzin

The series of four CASP experiments has helped to transform the field of protein structure prediction. The state of the art in protein structure prediction has undoubtedly changed, but has there been progress over the years?

"Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!"

The Red Queen to Alice in Through the Looking-Glass *by Lewis Carroll*

Unlike many of the conferences in the past year, this meeting did not have '2000' in its name. The CASP4* organizers, John Moult (CARB, University of Maryland, USA), Tim Hubbard (Sanger Centre, Cambridge, UK), Krzysztof Fidelis and Adam Zemla (both of the Prediction Center, Lawrence Livermore National Laboratory, USA), did not need to follow the millennium year fashion to attract due attention. CASP4 was more than just a scientific conference. It was a convention of dedicated people who spent their summer working on the predictions of a new batch of protein structures and then gathered together to take one more step in the ongoing CASP experiment. This experiment, designed by John Moult and colleagues seven years ago, has been run every other year, in 1994, 1996, 1998 and 2000. The details of all CASP experiments are available online (*http://PredictionCenter.llnl.gov/*). The CASP4 proceedings will be published in a special issue of the journal *Proteins* later this year, where the proceedings of the first three CASP experiments have been published[1–3].

Description of the experiment

The main objective of CASP was to subject the available prediction methods to a blind test. The participants in the experiment were asked to predict a number of structures that were about to be determined by X-ray crystallography and NMR spectroscopy. The target protein sequences were solicited from experimentalists. The targets were divided into three different categories: comparative modeling, fold recognition and *ab initio* methods.

This division reflected the status of the field at the beginning of the experiment in 1994. At that time, it had been well recognized that the structure of a protein is fully determined by the sequence of its amino acid residues, but all earlier attempts to compute the structure *ab initio* using nothing else but the sequence and the laws of physics and chemistry had very limited if any success. However, there were other methods available that used additional information. The observation of homologous proteins having very similar structures allowed the comparative modeling of protein structure by sequence homology to known structure. The notion that Nature may operate with a limited repertoire of protein folds gave rise to the fold recognition methods searching for a correlation between a given sequence and a known fold. A substantial fraction of the target proteins were expected to have new protein folds thus providing the opportunity of a double blind test of the *ab initio* methods by eliminating possible insights from the fold recognition methods.

The boundaries between the categories were somewhat arbitrary and have changed from one CASP experiment to another. Also, the CASP process stimulated the development of new approaches to protein structure predictions, for example, the knowledge-based methods that could predict across the different categories. To make the assessment as fair as possible to all prediction methods, after CASP2, the prediction formats were unified for all categories, and the category of each target was decided upon the assessment of submitted predictions. The assessment was carried out by independent assessors invited to analyze the predictions in each category and to nominate the best predictors for the presentation of their methods at the final meeting.

The original CASP design has withstood the test of time. It has not changed significantly since the first CASP experiment, neither has the format, timing or location of the final CASP meetings. The experiment has been a success among the predictor community, each time attracting more and more participants and generating more and more predictions (Table 1). In contrast, target collection, which depends on the genorosity of experimentalists, failed again to reach an optimal total number of 100 targets in all categories. In CASP4, there were just 43 targets, the same number as in CASP3. On the positive side, 40 of the CASP4 targets had their experimental structures determined in time and thus were available for the assessment, four more than were available in CASP3.

Assessment of predictions

The assessors' burden of interpreting the results was heavier than ever this year. To deal with it, each of the three principal assessors requested the assistance of his/her colleagues. Anna Tramontano (IRBM, Pomezia, Italy), the co-organizer of two FEBS advanced courses on the Frontiers of Protein Structure Prediction in 1995 and 1997, was in charge of the comparative modeling assessment. Manfred Sippl (University of Salzburg, Austria), the leader of one of the most successful predictors teams of the three previous CASPs, carried out the fold recognition assessment with help of his group. The former CASP2 assessor in the *ab initio* methods category Arthur Lesk (University of Cambridge, UK) made a comeback as the principal assessor in the renamed category of new fold methods. The renaming from *ab initio* was prompted by the development of new prediction methods that assemble a protein fold from small parts using both the first principles and empirical rules derived from known structures. The assessors had just two months before the final meeting to complete their analyses of more than 11,000 predictions submitted by 163 predictor teams and automated servers. Although the assessment was facilitated by the numerical evaluation data generated in the Prediction Center and by the assessors' own software, it was partly manual labor and extremely time-consuming (M. Sippl estimated that the fold recognition assessment had required ~25 person-weeks working time). In the fold recognition and new fold methods categories, manual inspection was needed for the identification of partially correct predictions that do not stand out in the numerical evaluation tables and plots. In the comparative modeling category, where almost all pre-

Table 1 CASP process in numbers

CASP number	Year	Number of targets	Number of predictor teams	Total number of predictions
CASP1	1994	33	35	135
CASP2	1996	42	72	947
CASP3	1998	43	98	3,807
CASP4	2000	43	163	11,136

dicted structures were sufficiently close to the experimental structure, the focus of manual inspection was on the prediction of fine details.

The assessors also were asked to address the following question: has there been progress in comparison to the earlier CASPs? Two of the three principal assessors answered positively. The comparative modeling assessor was less certain, as there was no substantial difference in the quality of CASP4 and CASP3 models. But the same could probably be said about the predictions in the other two categories. At first glance the CASP4 and CASP3 results look quite similar. In both CASP3 and CASP4, there were many correct predictions for some targets, whereas for other targets there were just a few good predictions, and a few targets were missed completely.

The non-uniform distribution of successful predictions could be explained by the variation of prediction difficulty. With increasing prediction difficulty, both the number of correct predictions and the mean accuracy decrease. For each target, however, the prediction difficulty is subjective and depends on the prediction method. For example, the targets with many known structural homologs are considered to be easier for the fold recognition methods than targets with few similar known structures. Other specific factors include the sequence similarity in the structural alignment and the amount of common structure shared by the target protein and the protein of most similar known structure. For the sequence similarity based methods, the targets from large sequence families are generally easier than the targets from small families or with orphan sequences. The knowledge of target protein function can be of great help for the knowledge-based methods. Previous CASP experience also comes in to play; for example, in CASP4, many predictors readily recognized the structural relationships of several fold recognition targets to some previous CASP targets considered rather difficult at the time. This illustrates that the prediction difficulty cannot be measured on an absolute scale; therefore the difficulties of the CASP4 targets are not directly comparable to the difficulties of the previous CASP targets, neither are the statistical results of different CASPs.

Development in structure prediction

The CASP process has brought a strong element of competition to the field, in particular to the fold recognition category, where the number of correctly assigned folds is a simple criterion for a team's success. In the earlier CASPs, only a few prediction teams consistently made a substantial number of correct assignments. In CASP4, there were many more teams that showed similar good performances over a wide range of targets, so this time the assessors measured the team's success by quality rather than quantity of the team's correct predictions. In pursuit of the best result rather than the best method, the approaches used by different teams have begun to converge over the years. Many predictors used combinations of different techniques rather than a single method to improve their performances.

The CASP4 predictors were also able to benefit from the predictions submitted to the CAFASP2 (critical assessment of fully automated structure prediction) experiment[4], which was run in parallel with CASP4 on the same set of targets. The CAFASP participants were 33 fully automated servers and computer programs. The CAFASP metaserver automatically submitted the targets to each server and collected the server predictions within 48 hours after submission of the target. The server predictions were then made available online, providing a clear indication of the target difficulties and saving many human predictors from embarrassing mistakes on easy targets. The CAFASP2 predictions were assessed along with the CASP4 predictions. There were at least four servers that did quite well in CASP4, but the separate, manually submitted predictions by their developers clearly showed significant improvements by human intervention.

The fact that predictors have been learning from each other's experience and expertise was probably the most important factor that improved the protein structure prediction within the CASP process. This improvement, however, was not due to this factor alone. There were major developments outside the CASP process that transformed the field of protein structure prediction, particularly in the fold recognition category, arguably the most interesting in the field. Unlike the *ab initio*/new fold methods category, successful predictions in the fold recognition category could result in substantially complete structures; unlike comparative modeling, correct fold assignment was far from trivial. In the early days of CASP, the fold recognition category was dominated by methods that thread the sequence in question through a library of known protein folds, and the terms 'threading' and 'fold recognition' were synonymous. In CASP2, a major blow to the superiority of threading methods came from the knowledge-based methods brought up by the creation of databases of proteins of known structure, such as SCOP (the Structural Classification of Proteins database). The coming of multiple sequence alignment-based similarity searches, like PSI-BLAST, made a big impact in CASP3. PSI-BLAST allowed the confident detection of sequence homology between a target protein and a protein of known structure with sequence similarity much less than the pairwise method threshold of 30%, thus moving many of the would-be fold recognition targets into the comparative modeling category. This move affected only the subset of fold recognition targets, known as distant homology recognition targets. The distant homology recognition targets are not only structurally similar to the proteins of known folds, but also probably evolutionarily related to some of those proteins. The genome sequencing projects, which are revealing many novel sequences, will help further the advancement of multiple alignment methods. This process will eventually deplete the distant homology fold recognition targets and add them to the comparative modeling targets. It should be emphasized that in both CASP3 and CASP4 even the best fold recognition methods were successful mainly because of the distant homology recognition targets that were considered to be the easier ones. For the difficult targets, the fold recognition methods started to lose ground to the new fold methods. In CASP4, no new success came from the

threading methods. One of the top threading teams in CASP3 (that of David Jones, Brunel University, London, UK) fully automated their program Threader that was one of the top CAFASP2 servers, but skipped the manual post-processing of its results. The other two CASP3 leading threading teams did not predict in CASP4; Sippl's team carried out the assessment and therefore they were not allowed to submit their predictions, whereas Bryant's team (NCBI, NIH, USA) did not participate at all.

CASP4 highlights

The selection of main speakers showed that most of the CASP3 leading teams retained or, in some cases, strengthened their leading positions in CASP4. There were fewer speakers selected to present their results in detail in CASP4, allowing more predictors to make short presentations. At least four of the selected speakers did well in more than one category, but each of them was allowed only one main presentation.

David Baker (University of Washington, Seattle, USA) and his team performed outstandingly well across all three categories. In CASP3, their fragment assembly method produced several good predictions including arguably the best CASP3 prediction. In CASP4, this new fold method predicted complete folds of at least four new fold targets. Baker and colleagues successfully extended their approach to the prediction of new structural features in fold recognition and comparative modeling targets.

After serving as one of the CASP3 assessors, I teamed again with Alex Bateman (Sanger Centre, Cambridge, UK). Using essentially the same knowledge-based approach to distant homology recognition that performed best in CASP2, we produced the most accurate models for several fold recognition targets and achieved the top averaged score in this category. We also applied a knowledge-based approach to the prediction of new folds and had some successes.

Michael Sternberg's team (ICRF, London, UK) retained its leading position in the comparative modeling category (the talk was presented by Paul Bates) and improved on the prediction of fold recognition targets. The team's fold recognition server, 3D-PSSM, was the best of all CAFASP servers that participated in this category.

Leszek Rychlewski and Janusz Bujinicki (International Institute of Molecular and Cell Biology, Warsaw, Poland), the organizers of CAFASP-like LiveBench experiment[4], did well in comparative modeling and the prediction of distant homology targets. Their prediction strategies utilized sequence profile-profile methods and modern threading approaches, which were carefully benchmarked before in LiveBench.

Also, L. Rychlewski together with Arne Elofsson (Stockholm University, Sweden) and Daniel Fischer (Ben Gurion University, Beer-Sheva, Israel) compiled the consensus predictions by CAFASP servers. In the fold recognition category, CAFASP consensus (presented by D. Fischer) performed better than any single server, but, unlike other servers, reaching a consensus was not fully automated.

Other selected speakers included: in the comparative modeling category, Ceslovas Venclovas (Lawrence Livermore National Laboratory, Berkeley, California, USA); in the fold recognition category, Kevin Karplus (University of California, Santa Cruz, USA), Tom Blundell's team (University of Cambridge, Cambridge, UK, presented by Jiye Shi) and SB-fold (SmithKline Beecham Pharmaceuticals, Philadelphia, USA, presented by Andrej Lupas); in the new fold methods category, Jeffrey Skolnick (Danforth Plant Science Center, St. Louis, Missouri, USA), Richard Friesner (Columbia University, New York, USA), David Shortle (Johns Hopkins University, Baltimore, USA) and Rita Casadio (University of Bologna, Italy). The selected speakers presented new developments in their methods and/or found new areas of application for these methods.

Future challanges

The sustained success of many individual predictors demonstrates the progress made since earlier CASPs. The collective progress is less evident, due to rapid changes in and outside the CASP process. Like Alice in Wonderland who had to run as fast as she could to keep her place, the predictors must perform better each time just to keep their place in the CASP league table. The next big changes in the field almost certainly will come from structural genomics projects aimed at the experimental determination of a large number of novel protein structures.

The CASP process has to respond quickly to forthcoming changes to keep the prediction field going when structural genomics projects gain full speed. It can be anticipated that the role of fold recognition methods will eventually diminish, while the development of comparative modeling and new fold methods will get a new lease on life. Structural genomics is in a very good position now to help in speeding up the CASP process by providing more prediction targets. In return, CASP could probably help in the functional annotation of the target structures by sharing the gathered information that helped in making the correct predictions.

One of the ongoing structural genomics projects (*http://s2f.carb.nist.gov*) already provided several CASP4 targets. The structures of all these targets were successfully predicted, including two correctly predicted new folds that were among the CASP4 highlights. I would like to conclude with an appeal to structural biologists not to miss a good opportunity of fruitful collaboration between CASP and structural genomics.

Alexey G. Murzin is in the Centre for Protein Engineering, MRC Centre, Hills Road, Cambridge CB2 2QH, UK. email: agm@mrc-lmb.cam.ac.uk

*CASP4, the Fourth Meeting on the Critical Assessment of Techniques for Protein Structure Prediction, Asilomar, California, December 3-7, 2000.

1. *Proteins Struct. Func. Genet.* **23** (1995).
2. *Proteins Struct. Func. Genet.* **Suppl. 1** (1997).
3. *Proteins Struct. Func. Genet.* **Suppl. 3** (1999).
4. Fischer, D., Elofsson, A. & Rychlewski, L. *Protein Eng.* **13**, 667–669 (2000).