

# System for Accepting Server Predictions in CASP6

Volker A. Eyrich,<sup>1</sup> Andriy Kryshchak, <sup>2\*</sup> Maciej Milostan,<sup>2†</sup> and Krzysztof Fidelis<sup>2</sup>

<sup>1</sup>CUBIC, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York

<sup>2</sup>Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, California

**ABSTRACT** We describe the new CASP system for collecting and verifying predictions generated by servers. The system was developed to ensure reliable execution of the server assessment part of CASP, with particular emphasis on data consistency. Following the principle that predictions should not be modified by anyone but their authors and to allow a later meaningful assessment, submissions are now verified for correctness of format and contents within the strict 48 hour CASP deadlines for this type of submission. This article also provides an overview of the rules governing server participation in CASP6 and some statistics pertaining to servers in CASP6. *Proteins* 2005;Suppl 7:24–26. © 2005 Wiley-Liss, Inc.\*

**Key words:** protein structure prediction; server; CASP

## INTRODUCTION

Server predictions were paid special attention in CASP beginning in 1998<sup>a</sup> with the CAFASP initiative<sup>2</sup> (Critical Assessment of Fully Automated Structure Prediction). Since then we have been asking the same set of questions following each of the CASP/CAFASP experiments. How big is the gap in model quality between the best servers and the best human predictors? Is it closing and if so, how quickly? Which server methods look the most promising and which servers can generate the best models?

During three rounds of CASP (1998, 2000, and 2002), CASP had relied on CAFASP for collecting server predictions. In CASP6, however, due to the relatively tight deadline and format requirements, it became necessary to implement an independent system. This article gives an overview of the new system for accepting server predictions (Fig. 1) and of the rules used in the CASP6 server experiment. We also provide brief statistical data regarding server participation in CASP6. Assessment of CASP6 server predictions is not the subject of this article and is addressed in the assessors' articles in this issue.

## PARTICIPATION RULES

The CASP term "server" relates to a computer or a cluster of computers accessible through the Internet and capable of generating structural models of proteins from amino acid sequence without human intervention.

Participating CASP servers were expected to return their models within 48 hours from the time they received the query, provided that the load did not exceed three targets (approximately 700 residues on the average) per day. In CASP6, servers usually had slightly more time as the target sequences were posted on our website and sent out to servers in the morning of day *n* and server models were due by noon of day *n* + 2.

Participants were to have registered their servers for the experiment through the CASP6 website. Servers could participate in all CASP6 prediction categories. Server curators were also allowed to additionally participate as human-experts. The registered predictors were notified about the weekly target release schedule with an e-mail message several days in advance. Target queries were sent to servers directly from the CASP distribution server, and the returned, unaltered models were instantly processed by the CASP verification software. At the time of submission, each model was recorded in the system and assigned a PIN number. Then, all predictions were automatically checked for compliance with model format appropriate to their prediction category. Error or acceptance messages were automatically sent to the server curator via e-mail. No prediction was accepted until it passed format verification. If a prediction did not pass the verification system, the server curator had an additional 24 hours to correct the error in the submission format and resubmit directly to the CASP6 verification server through a separate gateway. Predictions that passed the format verification were considered accepted and assigned an accession number. After 72 hours, no server predictions were accepted or altered in any way.

Abbreviations of prediction categories: 3D, three-dimensional coordinates; DP, domain boundary; RR, residue–residue contacts; DR, disordered regions; FN, prediction of function.

Grant sponsor: U.S. Department of Energy; Grant number: W-7405-Eng-48 (to the University of California, Lawrence Livermore National Laboratory). Grant sponsor: National Institutes of Health; Grant number: LM007085 (to K. Fidelis).

†Current address; Institute of Computing Science, Poznan University of Technology, Poznan, Poland

\*Correspondence to: Andriy Kryshchak, Biosciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94551. E-mail: andriy@llnl.gov

Received 6 May 2005; Accepted 4 July 2005

Published online 26 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20719

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version.

<sup>a</sup>Secondary structure prediction servers were assessed for the first time in CASP in 1996.<sup>1</sup>

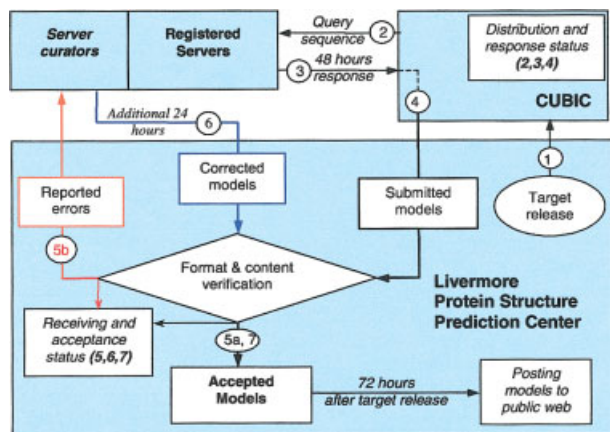


Fig. 1. Server CASP6 data flow.

## SYSTEM ORGANIZATION

### Registration of Servers and Dry Run

The server developer/curator registered a server using our Server Registration Form. If several people were involved in running a server, they could register as members of the server team. The server developer/curator chose a unique name for the server (CASP6 predictor group) and was automatically issued a registration code that could be used to check accepted models or to resubmit predictions if any discrepancies in the original submission existed (within the time limits described above).

Since the use of CASP format and the fact of immediate processing of server submissions by the CASP verification software were significant changes, we thought it prudent to start the server experiment with a “dry run.” Participating servers tested their ability to generate predictions in compliance with the required format, and we tested our system for connectivity and ability to process a high load of predictions. The CASP experiment started only after integration problems had been eliminated for the vast majority of servers.

### Distribution of Target Sequences to Servers

Target submission and prediction collection tasks were handled at CUBIC. We used the well-established server META<sup>3</sup> for submitting target sequences to servers. Predictions received were collected by a custom system built on top of EVA,<sup>4</sup> a system for automatic and continuous analysis of prediction servers. We used the procedure that follows.

At the time of the official release of a given CASP target, the respective sequence was submitted to the registered prediction servers via META using e-mail or direct connection to submission pages/CGI (Common Gateway Interface) scripts (depending on the participating server’s setup). The system made every attempt to distribute targets quickly and reliably in order to give participants as much time as possible. The obvious system problems on the remote server side were dealt with automatically whenever possible (e.g., targets were resubmitted in the case of connection failures, etc.). Nonobvious or nonrecoverable

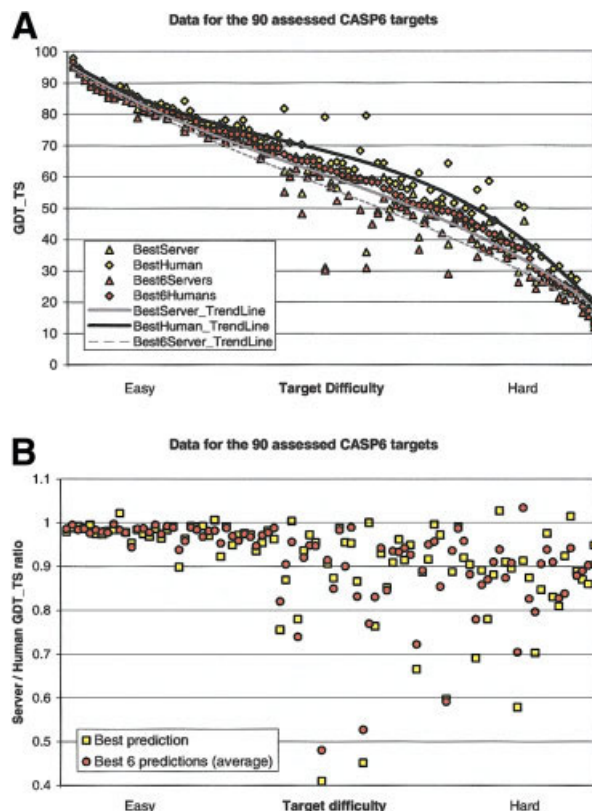


Fig. 2. Comparison of server and human groups performance in CASP6. Targets are sorted by their a posteriori predicting difficulty in terms of the average GDT\_TS score from the best six predictions (from different groups). (A) Data show that the best server models are almost as good as best human-expert models on the high homology targets and, on average, 12% worse on the more difficult targets. (B) Server/human GDT\_TS ratio; 69% of the best server predictions are within 10% of the best human predictions.

errors were dealt with manually (sometimes with assistance from the server administrators). Predictions returned by the servers were collected at CUBIC and immediately forwarded to Livermore Protein Structure Prediction Center for further processing.

To make the sequence distribution and prediction collection process more transparent CUBIC maintained status pages that provided detailed information on when exactly target sequences had been submitted, what the server response had been and when (or whether) prediction results had arrived at CUBIC.

All in all, the integration of all the CASP servers seems to have been relatively painless for predictors as well as CASP/CUBIC. In part this was probably due to the fact that most of the servers (or one of their predecessors) were already available via META; therefore, little extra integration work had to be done.

### Accepting Predictions

Predictions received at the Livermore Prediction Center were automatically checked for compliance with the CASP format, required to ensure proper subsequent evaluation. Our strongly held position is that after the predictions

leave the server, they should not be “touched” by anyone. To comply with this premise and make the transitional prediction season less stressful for server administrators, we relaxed the format requirements for server predictions whenever possible.

First, we allowed servers to return several models in one file. Our script split a multiple model file into separate files and only then sent each model separately to our verification software (models with numbers greater than five were ignored).

Second, to diminish CASP/CAFASP format incompatibility issues, we adjusted our scripts so that the CASP verification server was not rejecting predictions containing records required only by CAFASP. At the same time, it was able to obtain information required by CASP from records transparent to CAFASP but recognized by CASP.

Third, planning to make server models publicly available during the experiment, we asked server curators to modify the identification record in a prediction so as to ensure data access confidentiality.

Information about predictions received at Livermore was immediately available to the server predictors through the acceptance status page.<sup>5</sup>

### Public Availability of Server Predictions During the Experiment

In CASP5, many human-expert prediction groups used server output collected and publicly released by CAFASP as a starting point for their work, and the best models were sometimes generated in that way. To allow this approach in CASP6 we posted server predictions at our website immediately following the 72-hour prediction acceptance period. The Web page was organized so that one could navigate through the data starting from the server name or target ID as an entry point. Public availability of server results proved to be helpful both for human-expert groups (providing valuable information for manual modeling work) and also for servers (decreasing the unnecessary load from processing many identical queries).

### Server Results on the Web

Server predictions were assessed alongside the human-expert predictions using the same evaluation methods. Following the CASP6 meeting, server results were published online, together with the results from human-expert groups at the CASP6 website. In the results tables, the server groups are marked with an “s” in the group number field. It is possible to inspect server results separately by using the link “Server models.”

## SUMMARY STATISTICS

Participation was as follows:

- Sixty-two servers actively participated (submitted at least one prediction), including 50 servers participating in the 3D category, 11 in DP, 5 in RR, 2 in DR, and 2 in FN.
- There were 17,830 accepted server predictions, including 15,649 3D, 959 DP, 768 RR, 145 DR, and 309 FN predictions.
- There were 5309 server models designated as first that were accepted, including 3846 in 3D, 746 in DP, 421 in RR, 145 in DR, and 151 in FN category.

Below, we perform a simple, comparative analysis of human and server predictions across all CASP6 categories of 3D-assessment, aiming to provide an overall estimation of server predictions quality in CASP6. We want to emphasize that the statistics below are based on the GDT\_TS<sup>6</sup> (Global Distance Test Total Score) analysis only and, as far as GDT\_TS is not a perfect measure of accuracy, the results and plots below illustrate only general trends in prediction data:

- One hundred and two server-generated models were on the list comprising 540 top models from different groups (the best 6 models for each of the 90 domains assessed in CASP6).
- At least one server model was among the “best six” for 67% of domains assessed in CASP6.
- Seventy-two percent of server groups submitting 3D predictions produced at least one “best six” model.
- Best server predictions were at least as good as best human predictions for 7% of the targets, scored less than 10% below the best human predictions for 62% of the targets, and less than 20% below the best human predictions for 81% of the targets (details of this analysis can be found in Fig. 2).

## REFERENCES

1. Zemla A, Venclovas C, Reinhardt A, Fidelis K, Hubbard T. Numerical criteria for the evaluation of ab initio predictions of protein structure. *Proteins* 1997;Suppl 1:140–150.
2. Fischer D, Barret C, Bryson K, Elofsson A, Godzik A, Jones D, Karplus KJ, Kelley LA, MacCallum RM, Pawowski K, Rost B, Rychlewski L, Sternberg M. CAFASP-1: critical assessment of fully automated structure prediction methods. *Proteins* 1999;Suppl 3:209–217.
3. Eylich VA, Rost B. META-PP: single interface to crucial prediction servers. *Nucleic Acids Res* 2003;31:3308–3310.
4. Koh IY, Eylich VA, Marti-Renom MA, Przybylski D, Madhusudhan MS, Eswar N, Grana O, Pazos F, Valencia A, Sali A, Rost B. EVA: evaluation of protein structure prediction servers. *Nucleic Acids Res* 2003;31:3311–3315.
5. Available online at [http://www2.predictioncenter.org/Serv/menu\\_frames.html](http://www2.predictioncenter.org/Serv/menu_frames.html)
6. Zemla A, Venclovas, Moulton J, Fidelis K. Processing and evaluation of predictions in CASP4. *Proteins* 2001;Suppl 5:13–21.