heavily penalized, since in this case chances of having models in common with other groups are lower. One example: in Sali and colleagues' scheme, group 526 achieves a higher ranking than group 384. Both groups predicted four targets. The first selected very "popular" ones and obtained results comparable to the average, and the second achieved outstanding results on a set of targets that a significant fraction of the predicting groups (up to 40%) decided not to tackle and that were very difficult to predict accurately.

These criticisms of the Sali and coworkers scheme should not be interpreted as complacency on our part. We do recognize that there is much room for improvement in the CASP criteria. In particular, we appreciate the point that it would be useful to attempt to attach statistical significance to all of the CASP rankings. Much time and energy is wasted in arguing over the significance of rankings, distracting from the more important aspects of the results. A reliable way of assigning significance might ameliorate these difficulties, and be fairer to some predictors.

There is one final point where we do agree with the Sali and coworkers position. There is no doubt that large-scale bench marking, such as LiveBench [8] and EVA (http://cubic.bioc.columbia.edu/eva) will play an increasing role in the assessment of structure modeling methods. The original form of the CASP experiment was designed in 1993, and was tailored to conditions that existed then. In particular, the process is built around the concept of collecting targets from the experimental community within a fixed time window. There are now many new opportunities for performing effective measurements of performance of structure prediction methods. LiveBench and EVA were both discussed at the CASP4 meeting, and we expect more emphasis on these in CASP5. The CAFASP series of experiments [9], run in close collaboration with CASP, are providing an evaluation of automatic prediction methods. We hope that this will also be a feature of CASP5.

Other changes in the prediction assessment field are underway. The sequence information for some of the proteins "on hold" in the Protein Data Bank is now public. There is an international agreement to provide information on proteins under study in structural genomics projects, including progress in solving each structure (http://www.nigms.nih.gov/news/reports/airlie_tasks.html). A "model database" equivalent to the Protein Data Bank, but for computational models, will likely be established. These are all valuable new sources of prediction targets. New web-based services providing standardized evaluation of methods performance (http://predictioncenter.llnl.gov/local/ace/ace.html) [10] will also have an impact. We look forward to the further evolution of the CASP framework to incorporate these developments.

John Moult,[1,5] Krzysztof Fidelis,[2] Adam Zemla,[2] Tim Hubbard,[3] and Anna Tramontano[4]
[1]Center for Advanced Research in Biotechnology
University of Maryland Biotechnology Institute
Rockville, Maryland 20850
[2]Biology and Biotechnology Research Program
Lawrence Livermore National Laboratory
Livermore, California 94551
[3]Sanger Centre
Wellcome Trust Genome Campus
Cambridgeshire, CB10 1SA
United Kingdom
[4]Department of Biochemical Sciences
"A. Rossi Fanelli"
University of Rome "La Sapienza"
P.le Aldo Moro 5
00185 Rome
Italy
[5]Correspondence: jmoult@tunc.org

Received: December 8, 2001
Revised: February 11, 2002
Accepted: February 11, 2002

References

1. Marti-Renom, M.A., Madhusudhan, M.S., Fiser, A., Rost, B., and Sali, A. (2002). Structure *10*, this issue, 435–440.
2. Moult, J., Fidelis, K., Zemla, A., and Hubbard, T. (2001). Proteins Suppl. *5*, 2–7.
3. Tramontano, A.L., Leplae, R., and Morea,V. (2001). Proteins Suppl. *5*, 22–38.
4. Lesk, A.M., Lo Conte, L., and Hubbard, T.J. (2001). Proteins Suppl. *5*, 98–118.
5. Sippl, M.J., Lackner, P., Domingues, F.S., Prlic, A., Malik, R., Andreeva, A., and Wiederstein, M. (2001). Proteins Suppl. *5*, 55–67.
6. Zemla, A., Venclovas, C., Moult, J., and Fidelis, K. (1999). Proteins Suppl. *3*, 22–29.
7. Venclovas, C., Zemla, A., Fidelis, K., and Moult, J. (1997). Proteins Suppl. *1*, 7–13.
8. Bujnicki, J.M., Elofsson, A., Fischer, D., and Rychlewski, L. (2001). Protein Sci. *10*, 352–361.
9. Fischer, D., Elofsson, A., Rychlewski, L., Pazos, F., Valencia, A., Rost, B., Ortiz, A.R., and Dunbrack, R.L., Jr. (2001). Proteins Suppl. *5*, 171–183.
10. Leplae, R.H., and Hubbard, T.J.P. (2002). Bioinformatics, in press.

PII S0969-2126(02)00729-3

Reply to Moult et al.

Here we address in brief several criticisms of our paper offered by Moult et al. We do not bring up the many points of agreement already mentioned by Moult et al., although they are greatly appreciated.

In our short paper, we did not aim to analyze the state of comparative modeling, nor to propose specific criteria for assessing the accuracy of comparative modeling. Instead, we described how to assess the statistical significance of ranking of methods given a model quality criterion. Applying the ranking method and one particular model quality criterion that has in fact been used previously, we illustrated difficulties with ranking of comparative modeling methods at CASP4. We suggested that CASP use some measures of statistical significance for whatever model quality criteria are adopted.

Their criticism of our model quality criterion appears to be missing these points.

We neither state nor imply in our paper that "the production of an unambiguous ranking of all modeling groups" is a major goal of the CASP experiment. However, it is worth mentioning that a specific ranking is a dominant outcome of CASP, despite the organizers' declared intentions.

Moult et al. suggest that if our scheme were applied at CASP, unreasonable ranking would ensue in some cases, such as the ranking of method 526 higher than method 384. However, we do not rank the two methods with respect to each other (Figure 2), since they share less than two models in common. The two methods are simply not comparable. This example actually reinforces our point that ranking of methods requires a sufficient number of common models.

We do not "insist that we must have more data so as to identify the small differences in the adeptness of the different groups" at CASP. We repeatedly stated that either a larger number of models, a larger average model quality difference, or a smaller standard deviation in the model quality difference are needed for reliable ranking. We showed that the current number of models is insufficient to make strong statements about the ranking of the methods at CASP4 given their differences. We also suggested that a larger number of models is required because modeling methods need to be assessed by a variety of criteria (see the next paragraph).

While Moult et al. agree with us that there are not enough models at CASP to assess comparative methods with statistical significance, they nevertheless claim that the field is "stuck". Even though we have not addressed ranking of methods applied to different small sets of targets at different CASP meetings, we believe it is not possible to say that there are no differences in accuracy between the comparative modeling methods applied at CASP2–4. For example, there have been only a handful of models at each CASP meeting for which the modeling of loops and side chains is not overwhelmed by the alignment errors. Thus, it is almost certainly impossible to make any meaningful statements about loop modeling and side chain modeling based on the CASP data. Outside of CASP, loop modeling and side chain modeling are the limiting factors in a number of important comparative modeling applications. As a consequence, there has not been "plenty enough CASP data to identify" major advances in comparative modeling.

We hope that our analysis helps both developers and users objectively assess the many existing protein structure prediction methods.

Andrej Sali,[1,3] Marc A. Marti-Renom,[1]
M.S. Madhusudhan,[1]
András Fiser,[1] and Burkhard Rost[2]
[1]Laboratories of Molecular Biophysics
Pels Family Center for Biochemistry
and Structural Biology
The Rockefeller University
New York, New York 10021
[2]CUBIC
Department of Biochemistry
and Molecular Biophysics
Columbia University
New York, New York 10032

[3] Correspondence: sali@rockefeller.edu

Received: January 4, 2002
Revised: February 14, 2002
Accepted: February 14, 2002