DOE/ER/62271-1

INFRASTRUCTURE FOR COLLABORATIVE PROTEIN STRUCTURE PREDICTION

Project Period FROM: 09/15/1996 THRU 8/31/2002

CARB/UMBI Principal Investigator: John Moult

University of Maryland Biotechnology Institute
Baltimore, MD 21202

Prepared for

THE U.S. DEPARTMENT OF ENERGY
AWARD NO. DE-FG02-96ER62271-M004

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Summary

This project had an unusual form: it was designed to promote comparison and exchange of results within the protein structure prediction community. The main activities were centered around web sites and workshops, and the principle challenge was to involve as many of the relevant members of the community as possible. This report covers (1) The CASP structure prediction experiments (CASP2, 1997 and CASP3, 1998 , CASP4 2000, each partially supported by this grant), (2) A community wide experiment to compare potentials for protein structure prediction (termed PROSTAR), and (3) Development of web based systems for the exchange of information on structure prediction methods. In our view, the CASP component of the project has been and continues to be very successful, with participation by almost all members of the protein structure modeling community, and substantial progress in the field to a large extent the result of CASP.

This report is written some time after the end of the funding period. Funding ended because DOE abruptly terminated the external computational biology program. Because of CASP's success, it has been possible to continue it with funds from elsewhere, in spite of withdrawal of DOE support, both for this project, and for the Protein Structure Prediction Center, until recently located at Lawrence Livermore. PROSTAR was also successful during the funding period, but has died since, for lack of support. However, much of the functionality and experience was rolled into another resource, 'Decoys are Us', run by Ram Samudrala, formally a graduate student of the Pi's.

Two other aspects of the work, comparison of conformational search methods and development of a web exchange system for function assignment to orphan genes, using structure prediction methods, were in progress at the end of the funding period. Function assignment of orphan genes has ceased. Comparison of conformational search has recently been revived, and is the subject of on ongoing experiment in the context of CASP (CASPR), with other funding.

Most of this report provides details of the CASP experiments, the most successful and lasting part of the work.

Introduction

Traditional funding of the development of structure prediction methods has focused on the individual investigator with a good idea. Our present state of understanding of the problem has grown out of that mechanism, and there will always be a need for a high level of such support. It is becoming increasingly clear, however, that complete solutions to the problem will require extensive, detailed development of these initial concepts. Further, such development is not possible within our present research infrastructure. One major example: It is critical to have reliable ways of distinguishing a correct structure from an incorrect one. Many groups have developed potential functions for assessing the correctness of a predicted conformation. At present there is no mechanism by which these functions may be extensively tested, and the results compared. Thus, we do not know how near to solved this problem is, or where further effort can most effectively be directed.

CASP is the most established community wide experiment so far. It has often been called a 'competition'. Inevitably, there are strong competitive aspects to it, and much time is wasted over arguments as to who 'won'. But the real strength of these processes is in the collaborative aspects: Most people in the community agree to test their methods on the same examples, and to have the results openly available. As a consequence, we all learn the strengths and weaknesses of every method there is, we all argue and think intensely about why some things appear to work and others not. And thus we all have at our disposal knowledge and insight on which to design our next round of algorithms.

Before the advent of fast, universal and almost free electronic communications. it would not have been possible to carry out this project. Essentially all of the process is electronic. While it is easy to exaggerate the impact that net technology will have on science, it clearly does open up a new means of effectively collaborating on a community wide level.

1). Critical Assessment of Protein Structure Prediction Methods (CASP)

(This section of the report is abstracted from a published article describing the CASP4 experiment, the last under this funding. Differences to earlier CASPs are noted). In CASP, methods of modeling protein structure are assessed on the basis of the analysis of a large number of blind predictions of protein structure. Early experiments focused on establishing what then current methods could or could not deliver. The later experiments have extended the significance of the earlier results by including a larger number of predictions from more investigators, and by measuring the extent to which there has been progress in each of the prediction areas in the intervening two years. With a series of four experiments conducted over a period of eight years, it has become clear where the bottlenecks to progress are, where progress is being made, and at what sort of rate.

As in CASP3, there was also a parallel experiment, CAFASP. CAFASP2 makes use of the CASP target distribution and prediction collection infrastructure, but is otherwise independent. The goal of CAFASP is to assess the state of the art in automatic methods of structure prediction.

A change in CASP4 was the inclusion of large scale benchmarking of prediction methods, EVA and Livebench. These experiments complement CASP, particularly by clarifying issues of the statistical significance of the results. Both operate by sending just released PDB entries to automatic prediction servers, and collating and analyzing the results over time.

The CASP4 Experiment

The structure of the experiment was very similar to that of the earlier ones, and consisted of three steps:
1. Information about 'soon to be solved' structures was collected from the experimental community and passed on to the prediction community.
2. Prediction teams deposited models of the structures before the experimental results were public.
3. The models were carefully compared with experiment, and a meeting was held to discuss the significance of the results.

Collection of Targets

X-ray crystallographers and NMR spectroscopists were solicited to provide information about structures that were either expected to be solved shortly or that were already solved but had not yet been discussed in public. Target information was made available to predictors through a web interface. Details of 43 structures were obtained, and of these 40 were solved in time to be included as official targets. A number of these targets were divided into two or more domains for assessment purposes.

Assessment

All CASP experiments so far have placed the primary responsibility for assessing the significance of the results in the hands of independent assessors. Papers by each of the assessment teams are included in the issue and constitute the most thorough and authoritative analysis available. As usual, the identities of the prediction teams were not known to assessors until they had completed an analysis and ranking of the results.

Meeting, Web site and Publications

Two planning meetings involving the assessment teams and the organizers were held at the Sanger Centre, one before assessment of the predictions, and one when a first round of prediction was complete. The meeting to discuss the outcome of the experiment was held at the customary place, in Asilomar, California in December, 2000. The assessors selected those prediction teams they considered had done the most significant work to talk at the meeting, and also those invited to write prediction report papers. Both at the meeting and in the papers, participants have been urged to concentrate on what went right, what went wrong, and where possible, to explain why, and what they learned as a result. All the papers in this issue have been peer reviewed. The CASP web site (http://predictioncenter.llnl.gov) provides extensive details of the targets, the predictions, and the numerical analyses

Progress over the CASPs

One of the main objectives of CASP is to measure progress in structure prediction. Between CASP1 and CASP2, there was a detectable improvement in prediction quality in many areas. In retrospect, it seems a large component of that may have been the community adjusting to the nature of the experiment. Since then, some areas have continued to advance, but not as rapidly, and some areas have remained almost static. The assessors papers address specific advancement areas, and the more general progress paper adds additional information. In summary, the picture is as follows: First the good news. The quality of new fold predictions has improved with every CASP, albeit starting from a very low threshold. In CASP4, for the first time, some contact predictions are approaching what may be a useful level of accuracy. The over-all success rate in detecting homologous folds relationships has also improved, with the best groups now able to identify a large fraction of the folds in this category. The not so good news: Although new fold models continue to improve, the best models for most targets are still not accurate enough to be useful for assigning function. Fold recognition does not work to a significant extent for analogous folds. Throughout the comparative modeling and fold recognition regimes, the accuracy of alignments remains the key limitation on the quality and usefulness of the models, and has hardly improved since CASP2. Comparative models are still at best no more accurate than can be achieved by simply copying the appropriate regions of template structures. Secondary structure accuracy may have improved slightly, but the improvement is at best small.

Subsequent Developments

At the time of writing of this report, there have been two more CASP experiments, in 2002 and 2004, conducted without the help of DOE funding. The size of the experiments has continued to grow with over 200 groups from 24 countries taking part in CASP6, and over 90 prediction targets, resulting in 40,000 predictions.

Members of the Organizing Committees

John Moult CARB, University of Maryland Biotechnology Institute, USA
Tim Hubbard Sanger Institute, Hinxton, UK
Jan Pedersen Acadia Pharmaceuticals, Denmark
Krzysztof Fidelis UC Davis, USA
Adam Zemla Lawrence Livermore laboratory, USA

2) Evaluation of Potentials for distinguishing between correct and incorrect protein structures.

Any method that aspires to predict aspects of protein three dimensional structure can only succeed if it employs some form of discriminatory function that is capable of reliably assessing the correctness of any conformation encountered. Although many such discriminatory functions have been developed, there has so far been no attempt to test any of them on a wide range of structure selection problems, or to compare the strengths and weaknesses of different functions. The potentials test set is intended to facilitate and encourage that type of testing. The goal of this sub-project is to allow as rigorous as possible a comparison of existing discriminatory functions so as to see what has been accomplished so far, and where future effort may most effectively be expended.

Results from the experiment are given below. It has the following components:

A. Collection of a set of decoys for testing the ability of a discriminatory function to detect the difference between correct and incorrect structures. Since different potentials are effective in different contexts, it necessary to collect and organize a variety of decoy types. The test set that has been established consists of the following decoy sets:

(i) Mis-threads: Sequences mounted on correct and incorrect folds (Holm and Sander, 1995). These decoys are appropriate for determining the ability of a discriminatory function to perform well in a threading application.

(ii) Incorrect PDB structures: Structures for which incorrect versions were initially deposited in the Protein data back. (Assembled with the aid of the obsolete PDB site, UCSF). These decoys are appropriate for testing the ability of a discriminatory function to detect particular types of errors, such as inadequate hydrogen bonding. Functions effective against these decoys may be used to validate experimental structures.

(iii) Correct and incorrect conformations of loops. (Fidelis et al, 1995). These decoys typically contain thousands of different conformations for very short stretches of chain. They are appropriate for testing the ability of discriminatory functions to correctly identify the fine details of a structure. (iv) Context independent protein fragments (Pedersen and Moult, 1997). These fragments are 12 to 20 residues long, and adopt conformations largely independent of the environment. They are appropriate for testing the ability of discriminatory functions to correctly identify supersecondary motifs.

(v) A set of conformations from a Monte Carlo trajectory of the folding of a small protein, protein G (Pedersen and Moult, unpublished). This is a range of conformations, starting from an extended chain, and finishing with a set of native like structures. They are appropriate for testing the ability of a discriminatory function to select the most compact and electrostatically satisfactory structure out of a large set of possibilities.

(vi) Molecular dynamics generated conformations of a medium sized protein, SGPA (Kitson et al, 1993). These conformations are all with about 2 Angstroms RMSD on all atoms from the experimental structure, and represent simple distortions. They are appropriate for testing the ability of a discriminatory function to get extremely fine details correct.

(G) A set of homologous models of protein structure, from CASP1 (PROTEINS, 1995). These models range from 1 to 4 Angstroms RMSD from the corresponding experimental structure, and contain incorrect features such secondary structures displaced as rigid bodies, incorrectly packed cores, and incorrect loop conformations. They are appropriate for testing the usefulness of a discriminatory function for comparative modeling.

B. Testing of the performance of discriminatory functions. The key component: how well do current discriminatory functions perform against the decoy sets?? Results for two of our own discriminatory function are public on the web, and a paper describing the tests has been published (Samudrala and Moult, JMB 275:895-916 1998).

C. Subsequent Developments

After the end of DOE support, PROSTAR has not been further developed. However, the student who worked on the project subsequently developed another site 'Decoys are us', which has been extensively used by members of the structure prediction community.

3) Development of a web based general system of exchange of information.

This component of the project was less successful during the granting period. However, based partly on experience gained then, we have since introduced a CASP discussion site (FORCASP), as an experiment information exchange.

Special issues of PROTEINS, reporting the CASP experiments:

PROTEINS **Suppl. 1**, 1997
PROTEINS **Suppl. 3** 1999
PROTEINS **Suppl. 5** 2001

Specific Publications (most in the special issues):

Criteria for Evaluating Protein Structures derived from Comparative Modeling:
Venclovas, A.Zemla, K.Fidelis & J.Moult
PROTEINS **29S** 7-13 1997

Critical Assessment of Methods of Protein Structure Prediction (CASP) - Round II:.
J.Moult, T.Hubbard, S.H.Bryant, K.Fidelis,& J.T.Pedersen
PROTEINS **29S** 2-6 1997

Critical Assessment of Methods of Protein Structure Prediction (CASP): Round III
J.Moult, T.Hubbard, K.Fidelis & J.T.Pedersen
PROTEINS **S3** 2 - 6 1999.

Processing and Analysis of CASP3 Protein Structure Predictions
C.Venclovas, A.Zemla, J.Moult & K.Fidelis
PROTEINS **S3** 22 -29 1999

Some Measures of Comparative Performance in the Three CASPs
C.Venclovas, A.Zemla, K.Fidelis & J.Moult
PROTEINS **S3** 231 - 237 1999

Predicting Protein Three Dimensional Structure
J.Moult
Curr.Opin.Biotechnol. **10** 583-588 1999

Critical Assessment of Methods of Protein Structure Prediction (CASP) - Round IV.
J.Moult, K.Fidelis, A.Zemla & T.Hubbard
PROTEINS **Suppl 5** 2-7 2001.

Comparison of Performance in Successive CASP experiments.
C.Z.Venclovas, K.Fidelis & J. Moult
PROTEINS **Suppl 5** 163-170 2001.

Processing and Evaluation of the Predictions in CASP4.
A.Zemla, C.Venclovas, J.Moult & K.Fidelis
PROTEINS **Suppl 5** 13-21 2001.