Who solved the protein folding problem?
Manfred J Sippl

For the third time, techniques for the prediction of three-dimensional structures of proteins were critically assessed in a worldwide blind test. Steady progress is undeniable. How did this happen and what are the implications?

Address: Center for Applied Molecular Engineering, Institute for Chemistry and Biochemistry, University of Salzburg, Jakob Hanninger Strasse 3, A-5020 Salzburg, Austria.

E-mail: sippl@came.sbg.ac.at

Structure April 1999, 7:R81–R83
http://biomednet.com/elecnet/0969212600700R81
© Elsevier Science Ltd ISSN 0969-2126

In December 1998, a crowd of predictors, assessors and onlookers assembled in Asilomar, California to quarrel over the current state of affairs in protein structure prediction. One hundred teams sacrificed their 98 summer vacation to compute a total of 3800 structural models for 43 protein sequences, the structures of which were determined in parallel by X-ray crystallography and nuclear magnetic resonance (NMR) laboratories. The heat continued in December when predictions were inspected, dissected and searched for traces of structural similarity to target folds and when teams and techniques were judged by rigorous assessors.

The details of this third critical assessment of techniques for protein structure prediction, or CASP for short, fill a forthcoming special issue of Proteins: Structure, Function and Genetics and summaries and comments on CASP3 appear in many journals. There is no need to repeat these observations here, and so we are free to dedicate this precious space to a reflection on protein structure prediction, the CASP idea, its life cycle, an overall view of the results obtained and their significance for structural biology.

The protein folding problem fascinates the scientist, the educated layman and the entrepreneur. For some, the folding protein is the most incredible message they ever saw — a message that turns itself into the thing it describes. For others, it is the philosopher’s stone that transforms cryptic sequences into fortunes. Whatever the reason, considerable efforts have been spent on cracking the folding code and the past 30 years are filled with success stories. Nevertheless, for a long time no practical solution has emerged.

The essential goal of protein structure theory is the computational determination, or ‘prediction’, of protein structures from amino acid sequences ahead of experiment. To demonstrate the value of a method the correct answer must be unknown in the calculation, but has to be available for evaluation. A serious researcher will exert self-control and simulate this situation, but in practice this is hard to achieve. Methods are often knowledge-based, relying to some degree on experimental data, and although a particular simulation environment might be accepted by some as unbiased, it could be rejected as biased by others.

Another issue is how to measure success in structure prediction. The targets are experimental structures determined by X-ray analysis and nowadays also by NMR. Neglecting variations in resolution and quality as well as problems of flexibility and dynamics, these targets are the ‘gold standard’. But then, how far is the arrow from the bull’s eye? This depends on the scale. At one extreme there is the pedantic crystallographer who uses his tiny ruler, counting small deviations as a failure. At the other extreme there is the generous theoretician who celebrates a globular blob as a big success. Hence, the statement “the protein folding problem is solved” holds a different meaning to different people.

To make progress, processes of self-control, common language and cooperation are required. For quite some time structural biologists were aware of this, but there was no solution. We do not know the date nor the exact circumstances of the CASP inspiration, but we do know that it was perceived by John Moult. To have an idea and to make it happen are quite different issues, however.

First of all, CASP needs suitable targets. The structures must be unknown to predictors, but available for evaluation. This unique constellation is found in X-ray and NMR laboratories, but the people there are not particularly inclined to talk about their unfinished projects and they usually do not see any advantage in a commitment to release their structures before they might be published.

Secondly, CASP needs predictors. These are serious scientists. They work on frightening formulae to compute complicated structures — and they know quite well how difficult it is to predict a protein fold. Why should they risk a likely failure? Finally, CASP needs assessors. An assessor cannot also be a predictor and her/his duty is to judge and criticize their colleagues work in public, a rather unpleasant and dangerous job. This sounds difficult enough, but on top of that CASP has to be organized as a worldwide experiment, and its ideas, results and conclusions have to be communicated to the interested community as well as
to the indifferent funding agency. To make CASP happen was not a small feat, but John Moult did it [1].

The CASP experiment is now a recurrent phenomenon in computational structural biology that controls the clock and mind of many workers in the field. The CASP cycle spans two years. The first year is used for recovery from the previous cycle, followed by a period of new inventions and developments. This quiet and often leisurely time ends when the prediction season starts in the second year, with the first targets appearing in spring.

As a general rule, activity on early targets is low. Possible explanations are that predictors are taken by surprise not having their gear ready, or that the lifetime of early targets is unusually short. In any case, the major flocks appear in summer with peaks in July and August, triggering a flood of submissions with a maximum peak during the closing days in early September.

Then follows the assessment, an uneasy period for predictors with tensions running high. This is a time of doubts, rumors, and self-made CASP winners. Rumors cease when the assessors choose the 18 groups whose predictions are considered to be the most successful or whose techniques seem to be the most interesting. The cycle ends with the meeting in December where assessors summarize their conclusions in Asilomar’s wooden chapel, along with the chosen ones who present their results and reflect on the themes ‘What went right?’ and ‘What went wrong?’

So much for the suffering, but what about happiness and fun in CASP3? There was, in fact, a lot — not counting the luring Asilomar beach and neglecting the bar and free beverages. Assessors and predictors presented exciting results with the intensity of emotions depending on your profession. We skip heavy numbers and minute details here to summarize a few essential points.

The 43 targets split into two groups. The first group contains targets that have detectable sequence homology to proteins in the Brookhaven Protein Data Bank (PDB). These are the targets for comparative modeling. For the predictor the goal here is to align the target sequence to the template and to model loops and sidechains. The second group consists of those targets that have no detectable sequence homolog in the PDB. For these targets there might be a related fold in the PDB, but who knows? The goal of the fold recognition teams is to spot such folds and to align the target sequence in a reasonable way. At the same time this second group of targets is the playground for those who start ab initio.

There were 20 targets comprising 22 domains in the fold recognition ab initio categories and as it turned out only two of these were new folds. One has to take this statement with a grain of salt, as structural similarity is a tricky thing that is hard to nail down quantitatively. But these two folds really are new. The remaining 20 folds span a range of similarity with roughly half of them having clear structural relationships to something in the PDB and the other half comprising borderline cases.

This result is in itself quite interesting, and even more so in the light of structural genomics. Most folds were known but we didn’t know they were. One highlight of CASP3 was that the fold recognition teams and threaders did know quite often. Taken together they correctly assigned 11 domains to a close structural homolog in the PDB and another four for the borderline cases. They failed in five of the difficult cases and they missed one domain in a multidomain protein that had clear structural homology to a fold in the PDB. The quality of results are not spread evenly, but there were at least six groups that did quite well. This soothes the troubled soul of the fold recognizer who often has to endure the evil mutter that all he does is align sequences in a complicated way. In fact, for most of these CASP3 targets there was nothing to gain on the sequence level.

How did things fare in the ab initio section? The practitioners in this corner frequently start at some unreasonable point in conformation space, and with the help of clever rules they try to get close to the native conformation. They often hear whispers that their results are no better than random. Not so in CASP3. There was no doubt that the fold had moved in the right direction in several cases and, most importantly, some of the predictions were better than the best template found in the PDB.

This leaves us with comparative modeling. Don’t change the template was the lesson from CASP1 and CASP2, and this is still good advice. Another key to success is an appropriate alignment technique, which is the domain of threading. No big surprise then that an uprooted threader had to confess in the comparative modeling session.

To summarize, quite a lot went right in this third CASP experiment. But does it have an impact on structural biology outside the circle of predictors? Yes it has — and its impact is quite significant. Here is just one example. Structural genomics is determined to spend a billion dollars to solve the 10,000 protein structures that are thought to be required to cover fold space. The project will benefit from some coordinated action.

Fold recognition is the method of choice to avoid redundant structure determination. These techniques are not yet perfect, but the more successful variants are able to spot related folds with a success rate of better than 60%. This number is an average; in fact, the answers are quite reliable in some cases and less reliable in others. Hence,
before hunting for a new fold consult a threader. If he can help just 10% of the time this might save you a hundred million dollars in the long run.

Then finally, is the protein folding problem solved? The lucky theoretician trumpets an overblown yes, but the grumpy crystallographer does not care to listen. And, by the way John, thanks for the rides on your roller coaster!

References