2

The Assessment of Methods for Protein Structure Prediction

Anna Tramontano, Domenico Cozzetto, Alejandro Giorgetti, and Domenico Raimondo

Summary

Methods for protein structure prediction are flourishing and becoming widely available to both experimentalists and computational biologists. But, how good are they? What is their range of applicability and how can we know which method is better suited for the task at hand? These are the questions that this chapter tries to address, by describing automatic evaluation methods as well as the world-wide Critical Assessment of Techniques for Protein Structure Prediction (CASP) initiative and focusing on the specific problems of assessing the quality of a protein 3D model.

Key Words: Protein structure prediction; accuracy of protein structure models; CASP; structure prediction servers; metapredictors.

1. Introduction

Protein structure prediction is a field that has attracted enormous interest since the very beginning of protein structural biology. The first model of a protein was produced only about 10 years after the first protein structure was solved and at a time when only two protein structures were available ***(1)***. The model was a physical one (no molecular graphics available at the time), but it was a rather good one; it was later established that the root mean square deviation (rmsd) between the alpha carbons of the model and those of the subsequently determined experimental structure was around 1 Å, a result that would be considered interesting even today.

From: *Methods in Molecular Biology, vol. 413: Protein Structure Prediction, Second Edition*
Edited by: M. Zaki and C. Bystroff © Humana Press Inc., Totowa, NJ

As of today, hundreds of servers and tools are widely available for producing a structural model of the protein of interest. The model can then be used as a structural framework for designing further experiments, interpreting functional data, assigning its molecular function to the protein, or as a target for drug design or even as a tool for solving the experimental structure of the protein and more. However, the quality of a model dictates its possible applications, and therefore, the admittedly complex problem of assessing beforehand the quality of models produced by different methods is of outstanding interest.

The issue is obvious: if one produces a model of a protein of known structure, the suspicion might arise that, unwittingly, data extracted from that structure are used in some of the steps of the procedure, and therefore, it would not be correct to extrapolate the results obtained on a test set composed of proteins of known structure to proteins of as yet unknown structure. On the contrary, predicting the structure of a protein for which no structural experimental data are available does not allow the effectiveness of the method to be assessed in a reasonable and predictable time frame.

The solution is to predict a protein structure "just in time" that is soon before the experimental structure of the protein is made available or before any method had a chance of taking the structure of the protein into account for optimizing its parameters.

The former strategy is used by the Critical Assessment of Techniques for protein structure prediction (CASP) experiment ***(2)*** and the latter by automatic evaluation servers such as EValuation of Automatic protein structure prediction (EVA) ***(3)*** and Livebench ***(4)***.

We will describe these experiments, give some advice about how to make the best use of the data they produce, and discuss their problems and limitations.

2. Materials

The models submitted to each of the CASP experiments and data related to their evaluation are available at http://www.predictioncenter.org. A discussion forum about most of the issues discussed in this chapter can be found at http://www.forcasp.org.

The EVA and Livebench automatic evaluation servers make their data available at http://cubic.bioc.columbia.edu/eva/ and http://bioinfo.pl/meta/livebench.pl respectively.

3. Methods

Predicting the structure of a protein is both an intellectual challenge and a practical issue, especially in light of the recent genomics and structural

genomics efforts. The problem is far from being solved in general terms, but it can be addressed using several heuristic strategies. During evolution, proteins tend to preserve their structure. It is therefore possible to derive information about a protein structure on the basis of the structure of an evolutionarily related protein, which, in turn, can be identified by sequence analysis [comparative modeling (CM)] ***(5)***. Even when no sequence similarity between two proteins can be detected, they might share structural similarity. In this case, the problem is to correctly recognize the compatibility of the sequence of the target protein with a known fold [fold recognition (FR)] ***(6,7)***. Finally, a protein might share neither sequence nor structural similarity with any known protein [new fold (NF)], and the prediction of its structure has to rely on different approaches. In many cases, when an NF is discovered, it is observed that it is composed of common structural motifs at the fragment or super-secondary structural level. This prompted the development of methods, known under the name of "fragment-based" ***(8,9)***, which try and assemble fragments of proteins of known structure to reconstruct the complete structure of a target protein.

4. The Difficulty of Evaluating a Prediction

At first sight, it might seem that the evaluation of the correctness of a model is a straightforward task once the experimental structure is available, but matters are not so easy.

First of all, the problem of finding the optimal superposition between two structures, that is, the superposition that minimizes some "distance" measure, does not have a unique solution. The difference between two superimposed structures depends on the fraction of the structures that is superimposed ***(10)***. It is entirely possible that one region of a model is very similar to the corresponding region of the target protein but that the similarity is masked if the whole structure is taken into account in the structural superposition. In other words, there is a relationship between the quality of a structural superposition and the fraction of superimposed structure. The identification of well-predicted regions not only is an issue related to the evaluation of the model but also might have important biological implications if they correspond to, say, the active site of the protein.

Furthermore, the measure traditionally used to evaluate structural similarity, the rmsd, is a quadratic measure. It is defined as the square root of the squared differences between the coordinates of corresponding atoms, and therefore, it will weight more regions that are not well superimposed with respect to the rest. From a biological perspective, if a region of a protein is incorrectly predicted, do we really care by how much or would we rather just like to say that the

predicted and experimental regions are more far apart than it is acceptable to derive meaningful insights from the model? This implies that the number of atom pairs of the model and the structure that are within an acceptable distance threshold is probably a better measure for the task of protein structure prediction evaluation.

Proteins are not static objects, they have a dynamic behavior and some regions are more flexible than others. We need to make sure that our quality measure takes this into account and does not penalize a model if it does not reproduce correctly regions of the experimental structure that have significant experimental uncertainty.

Furthermore, proteins are often composed of domains, and an evolutionary relationship between two proteins can be limited to one of the domains and not to the overall protein sequence.

5. The CASP Experiment

In 1994, John Moult proposed a world-wide experiment named CASP ***(2)*** aimed at establishing the current state of the art in protein structure prediction, identifying what progress has been made, and highlighting where future effort may be most productively focused.

Experimental structural biologists who are about to solve a protein structure are asked to make the sequence of the protein available, together with a tentative date for the release of the final coordinates. In the past 13 years, structural genomics consortia have significantly contributed to the set of CASP targets.

Predictors produce and deposit models for these proteins (the CASP targets) before the structures are made available. Another experiment, synchronized with CASP and called CAFASP ***(4)***, has been testing publicly available servers on the same set of targets, providing a unique opportunity for evaluating how much human expert knowledge is important to obtain better models. Recently, this task has been taken over by CASP itself ***(11)***. For testing server predictions, sequences are automatically sent to participating servers, and the models received within a short time frame, 48 h, are collected and stored. These models are also made available to human predictors, who have more time at their disposal, to avoid duplication of efforts, because many human predictors make use of automatic server results in their model-building procedure.

Finally, a panel of three assessors compares the models with the structures as soon as they are available and tries to evaluate the quality of the models and to draw some conclusions about the state of the art of the different methods. The experiment is run blindly, that is, the assessors do not know who the predictors are until the very end of the experiment.

Each of the routes to the prediction of a protein structure described before has traditionally been mirrored by a CASP "category," evaluated by one of the three assessors. The categories have some degree of overlap: CM targets for which evolutionary relationships are very hard to identify before knowing their structure can also be considered in the FR category; NFs can share some similarity with existing folds and be considered in both FR and NF categories. Recently, some modifications have been proposed, and the target categories will be reduced to two: template based and non-template based; but a special analysis will be performed on the best models to evaluate the accuracy of details of protein structure predictions, such as positioning of side chains and correct prediction of loop structures. The reasons for this rearrangement will become clear later.

The results of the comparison between the models and the target structures are discussed in a meeting where assessors and predictors convene; the conclusions are made available to the whole scientific community through the World Wide Web and through the publication of a special issue of the journal "Proteins: Structure, Function, and Bioinformatics."

There are several other categories that have been introduced in CASP throughout the years, such as prediction of function, of domain boundaries and of disordered regions, but we will not discuss them here.

The CASP experiment has been extremely successful. It has been repeated every 2 years since its first edition, and there is no sign that it is going to be discontinued in the near future ***(12)***. It is a very important experiment, which has the merit of having raised the issue of objective evaluation of structure prediction methods, of prompting the development of the automatic assessment methods that will be described later and of fostering the development of similar initiatives in other fields such as the prediction of protein—protein interaction, gene finding, and scientific literature mining.

6. CASP Measures

As we mentioned, there are two problems with the measure of the similarity between a model and a protein structure: the dependence of the solution on the fraction of superimposed structure and the quadratic form of the rmsd. One solution to the first problem is to use a graph such as the one shown in **Fig. 1**, where the x-axis indicates the fraction of the model that has been superimposed to the target structure and the y-axis reports the corresponding rmsd value (or any other similarity measure) ***(13)***.

In the last edition of CASP, there were almost 30,000 submitted 3D models ***(14)***, and it is not possible for any assessor or user to visually inspect all the

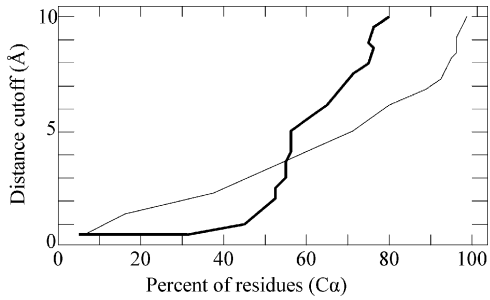

Fig. 1. A plot describing the quality of two predictions in CASP6 for target T0196 (an hypothetical protein from *Pyrococcus furiosus*, PDB code 1XE1). The *x*-axis indicates the percentage of aligned residues of the target and experimental structure that are closer than the threshold reported on the *y*-axis. As it can be seen from the plot, one of the models (indicated by the thick line) is closer to the experimental structure for about 60% of the structure, whereas the other turns out to be closer when larger fractions of the modeled and experimental structures are superimposed.

generated plots; so, it is necessary to convert the information into a numerical value, for example, a rough estimate of the area under the curve. The Global Distance Test (GDT-TS), used in CASP, is such a measure. It is defined as the average percentage of Cα atom pairs under a distance cutoff of 1, 2, 4, and 8 Å.

This measure is reasonably satisfactory for highlighting the overall quality of the prediction of the backbone of the protein, but it does not capture the details of the structure, for example, the correct prediction of the conformation of side chains. The latter is evaluated using the number of chi angle values within a threshold (usually set to 30°).

The next problem is related to the experimental uncertainty of the protein structure. CASP provides data for the complete model structure but also for subsets including, for example, all atoms that have a B-factor lower than a threshold (usually 20 Å), residues whose chi angles can be assigned reliably by X-ray crystallography, residues buried in the core, and so on.

Last but not least, CASP also analyzes the predictions of each domain of the target proteins separately.

7. The Problem of Evaluating the Overall Performance of a Method

The final aim of CASP is to highlight which methods work better, and therefore, it is essential to devise a comprehensive measure of the performance of a method on the basis of the results that the method achieved on several targets. And here, things get tricky.

First of all, not all methods are applied or applicable to all CASP targets, and therefore, a comparison between two methods needs to take into account how many models have been submitted using that method, but most importantly, which models. In fact, not all protein structures are equally difficult to predict, so that the relative difficulty of a target should be taken into account. The same problem arises when one wants to ask the obvious question of whether there has been any improvement of the methods in different editions of the experiment: each experiment has its own set of targets; therefore, the performance in one edition should be compared to the performance in another one taking into account the relative difficulties of the targets. The problem, as we will discuss in the **Subheading 8**, is a very complex one, but also extremely important for protein structure prediction evaluation.

8. Evaluating the Difficulty of a Prediction Target

The difficulty of predicting the structure of a given protein can be evaluated a posteriori, analyzing how well it has been predicted on average. In some cases, it is also possible to estimate the difficulty a priori. For example, in CM, one can see how difficult it is to identify the evolutionary relationship between the target protein and the protein of known structure that can be used as template for building the model and how easy it is to obtain a reasonable sequence alignment using standard methods. In FR predictions, one can measure how strong is the sequence-structure fitness signal. In both cases, one can also take into account, in evaluating the difficulty of modeling a protein, how well automatic methods perform the task.

It should be mentioned upfront that none of these strategies is faultless. For example, a posteriori evaluation cannot be used to compare two different CASP experiments, because, hopefully, methods have improved during the two intervening years, and the same is likely to be true for sequence alignment methods. Another effect, even more difficult to take into account, is the increased size of databases.

Traditionally, the difficulty of producing a comparative model for a protein has been measured on the basis of the percent of sequence identity or similarity between the target protein and the protein of known experimental structure used as template for modeling. However, although this measure takes into account the structural effect of the accumulation of mutations in the protein, it is not equally effective for estimating the difficulty of detecting the relationship and of obtaining a correct sequence alignment, that is, of detecting the right correspondence between the amino acids of the target and template proteins. In fact, most methods for the detection of sequence similarities rely on multiple

sequence alignment, that is, on information provided by many sequences of the proteins of the same evolutionary family. The increased size of the database can therefore be directly responsible for the improvement in the detection of evolutionary relationships and in the sequence alignment step, which are the essence of the quality of a model.

In CASP, the difficulty of a prediction is estimated on the basis of both its sequence and structural similarity with the potential templates. The former is defined as the fraction of structurally aligned residues (within 5 Å) that are identical between the target and the template, the second as the fraction of pairs of target–template Cα atoms within 5 Å after optimal superposition ***(15)***. When an 1D scale for target difficulty is needed, the average of the two values described above are used.

Another possibility is illustrated in **Fig. 2**. The multiple sequence alignment for each target available at the time of each experiment can be used to calculate the pair-wise sequence identity between each pair of sequences and to construct a graph similar to that shown in the figure. Each node represents one of the sequences in the multiple sequence alignment, and the lengths of the edges are proportional to the distance (inversely proportional to the percent of identity) between the connected nodes. The multiple sequence alignment is a path in the graph that includes all the sequences. In first approximation, the difficulty of aligning the target and template sequences depends on the availability of intermediate sequences, and this is determined by the most difficult pair-wise alignment that we need to perform to go from the target to the template. In other words, we might end up aligning a target and a template sequence only sharing

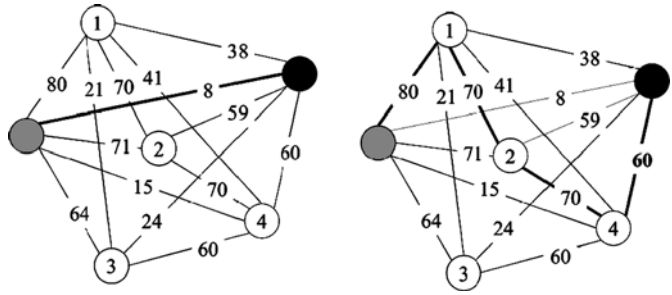

Fig. 2. Graph associated with a multiple sequence alignment containing a target (gray node) and a template (black node). Edges are weighted with the percent identity between the sequences they connect. Although the target and the template only share 8% of identical residues, the recruitment of homologous sequences allows to progressively align pairs of sequences sharing at least 60% sequence identity.

a very low sequence identity, but we might achieve this by aligning pairs of very similar intermediate sequences, starting from the target and "jumping" from one sequence to another until we reach the template much in the same way as we might cross a large river jumping from one emerging stone to the next. The difficulty of crossing the river is not proportional to its width but to the longest jump that we need to make.

Therefore, given all possible paths including target and template, we are interested in the one(s) where the maximum distance between each pairs of traversed nodes is minimal. Once such a path is found, the longest edge in the path, that is, the sequence similarity between the two most diverse sequences in the path is an estimate of the difficulty of aligning target and template, given the distribution of sequences in the multiple sequence alignment ***(16)***.

This approach gives, in first approximation, a measure of the difficulty of aligning the target and template sequence for each target in different experiments, given the database available at the time of the prediction, and can be used to ask whether the alignment of targets and templates of equivalent difficulty has become more accurate with time. **Figure 3** shows a plot of the percent of correctly aligned residues (a residue is considered correctly aligned if, after superposition of the experimental and modeled structure, its Cα atom falls within 3.8 Å of the corresponding experimental atom, and there is no other Cα atom of the experimental structure that is nearer) achieved in the last three

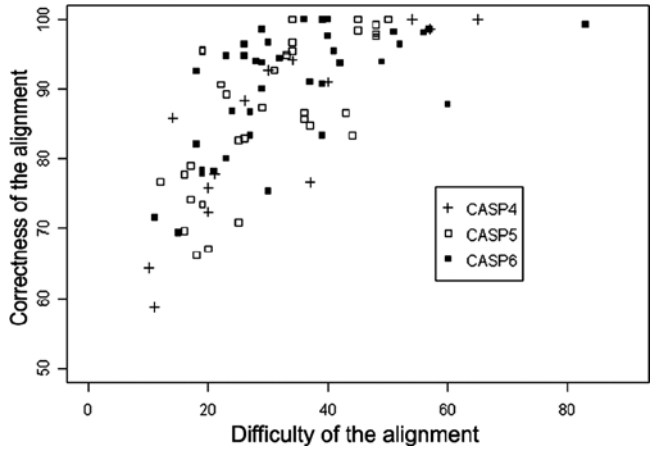

Fig. 3. Scatter plot of the alignment quality obtained in the last three editions of the CASP experiment as a function of the difficulty of the alignment, computed through the method depicted in **Fig. 2**.

CASP experiments for CM targets as a function of the difficulty parameter defined above.

As it can be seen, there has been no major improvement in methods for aligning sequences in the most recent CASP editions, and targets of similar difficulties are aligned with the same level of accuracy. This is somewhat disappointing and urges for novel ideas in the area.

Traditional methods for CM are based on the assumption that each of the modeling steps, including template selection, and alignment, can be optimized separately. It is easy to argue that a better approach would be to optimize all the parameters simultaneously. Clearly, this is beyond our present computational capabilities. However, it is worth noting that the most successful groups in recent CASP experiments used the strategy of constructing several models for each target protein and selecting the most likely one only at the end of the complete model-building procedure. In other words, rather than optimizing each of the steps of the comparative modeling procedure independently, they chose to also funnel sub- optimal intermediate results into each subsequent step. This represents a first degree approximation to a full multi-parameter optimization procedure, and we argue that this type of strategy should be pursued even more aggressively in the future.

It should also be mentioned, however, that predictors in CASP are not necessarily in an ideal position to produce the best models because of the time limitation imposed by the experiment. Also, the fact that the results are public and very visible might stop predictors from trying "risky" innovations.

9. New Challenges

There is no doubt that modeling methods are extremely powerful. At present, experimental structures are known for less than 1% of identified proteins, whereas relatively reliable models can be produced for up to 20% of proteins. In addition, models play an important part in a number of methods for obtaining structural data.

On the contrary, genomic efforts are producing the sequences of an impressive number of proteins, and there is no hope that all of them can be studied experimentally in the foreseeable future. Scientists do need to rely more and more on protein models to understand the function of this plethora of proteins, and, consequently, the required level of accuracy of a model, especially in the details of the structure, is increasing. CASP has highlighted a number of substantial improvements in modeling techniques, such as the development of FR and fragment-based methods, but, unfortunately, improvements in accurately predicting the details of a protein structure (such as positioning of

side chain and of structurally divergent regions, i.e., regions of the target protein that deviate substantially from the template) have not been equally satisfactory ***(12)***.

The overall conclusion that can be drawn from the analysis of the thousands of model submitted by hundreds of groups is that, rarely, a comparative model is closer to the experimental structure than the template used to build it or to reliably predict structural divergent regions. Furthermore, there seem to be no method able to consistently improve the accuracy of an initial model. An important goal is therefore to foster the development of modeling methods aimed at reaching an accuracy approaching the experimental error ***(17)***.

This is the rationale behind the emergence of a new category in CASP, aimed, as we mentioned, at evaluating the quality of the details of the models rather than their overall accuracy. It will be included in the next round of the experiment, and, hopefully, it will be as effective in pushing the field farther as the other CASP categories have been in the past.

10. Automatic Evaluation Servers

CASP is aimed at evaluating the state of the art in prediction methods; however, not all experimentalists interested in obtaining a model of their protein of interest have access to collaborations with outstanding modeling groups. The most common route to prediction for the majority of scientists relies on publicly available automatic servers. It is clearly important to evaluate the accuracy of these servers on a large set of data and in a continuous fashion.

This need has prompted the development of automatic systems that continuously evaluate automatic prediction methods. They collect the predictions returned by different servers for new protein structures before any method had a chance to use them in the training set.

EVA ***(3)*** is one of the servers that performs this useful service to the community. Every day, EVA downloads the newest protein structures from the Protein Data Bank (PDB) archive ***(18)***, extracts the sequences for every protein chain, and sends them to each prediction server registered for the experiment. The collected results are then evaluated and made public.

EVA covers several methods that predict solvent accessibility, secondary structure, and complete 3D modeling. The proteins used in the experiment are such that no pair of them has more than 33% identical residues over more than 100 residues aligned.

Another continuous benchmarking server is Livebench ***(19)*** that limits itself to the evaluation of 3D models of proteins not sharing a significant sequence similarity (and therefore deemed to be non-homologous) to any protein of

known structure. Every week, new entries in the PDB database with a length comprised between 100 and 500 residues are submitted to participating servers and their returned predictions collected and analyzed.

The results of both servers, together with some statistical evaluation of their significance, are publicly available through Internet, and they represent extremely useful tools that should be consulted before using any prediction server.

The possibility of automatically collecting the results of several prediction servers also prompted the development of the so-called metapredictors ***(20)***. These are gateways to various methods for protein structure prediction, which "outsource" the prediction task to publicly available servers, collect the results, and evaluate them. Some metapedictors just score the predictions and provide the user with a ranked list, whereas some others combine the predictions returning a single model. They usually perform better than single servers and probably represent the best solution to automatic prediction of protein structure as of today.

11. State of the Art of Structure Prediction Methods: The Usefulness of Protein Models

We said in the introduction that the quality of a model dictates its usefulness for several applications. As we discussed, estimating the quality of a model is not an easy task. However, some rules of thumb can still be provided, with the caveat that they are just indications and that each protein modeling experiment has a story of its own.

Comparative models built on the basis of a significant sequence identity between target and template, above 50–60% are certainly accurate in their overall structure and can be reliably used to analyze the conserved regions of the protein, such as its active site. As we mentioned, apart from special cases ***(21)***, the predictions of structurally divergent regions is likely of being much less accurate than the rest of the protein, and it is rather risky to derive biological conclusions from their conformation ***(22)***. For very high sequence identity, above 90%, there are usually very few structurally divergent regions, but here, the devil is in the positioning of the side chains. It has been shown that even models of high accuracy would fail if used as targets for drug design because the positioning of the side chain would not be sufficiently accurate ***(23)***.

For comparative models, a user should always take into account that the accuracy of the model is not uniform throughout the structure and that functionally important regions are likely to be better conserved, at least for orthologous proteins, than the rest of the structure. Comparative models based

on distant evolutionary relationships have been often instrumental in deriving functional properties of the protein, because these are usually brought about by the most conserved parts of the structure, which, in turn, are those predicted more accurately ***(24)***.

Models based on low sequence identity (below 30%), FR methods, and fragment-based methods should only be used as structural frameworks to think about the protein and certainly not for deriving detailed measures of distances or energies. Remember that, if the model is built by comparative modeling, we can at least be sure that the overall topology of the protein is correct, whereas this might or might not be true for fold recognition and fragment-based models. In these cases, only experimental verifications of the features predicted by the model can increase the confidence in a model.

Models can also be used for speeding up the experimental determination of a protein structure. For example, models with a GDT-TS value above 84 are consistently able to solve the phase problem in crystallography, that is, to be used as a tool to estimate the phases of the X-ray diffracted waves, a major problem in X-ray crystallography ***(25)***. Models can also be useful in speeding up the solution of the structure of proteins by nuclear magnetic resonance spectroscopy.

The impressive thrust of biological and computational methods makes it very difficult to predict what we can expect even in the near future. Nevertheless, more and more protein sequences and structures will become available, and there is no doubt that the sheer power of the data will help building more accurate protein structure models. On the contrary, if we look at the history of the past few years, we cannot but expect that new prediction methods will appear. It follows that the possibility of exploring the complete space of protein structure is, finally, within our reach.

References

1. Browne, W. J., North, A. C., Phillips, D. C., Brew, K., Vanaman, T. C. and Hill, R. L. (1969) A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme. *J Mol Biol*, **42,** 65–86.
2. Moult, J., Pedersen, J., Judson, R. and Fidelis, K. (1995) A large-scale experiment to assess protein structure prediction methods. *Proteins*, **23**(3), ii–v.
3. Koh, I. Y., Eyrich, V. A., Marti-Renom, M. A., Przybylski, D., Madhusudhan, M. S., Eswar, N., Grana, O., Pazos, F., Valencia, A., Sali, A., et al. (2003) EVA: evaluation of protein structure prediction servers. *Nucleic Acids Res*, **31**(13), 3311–3315.

4. Fischer, D., Elofsson, A. and Rychlewski, L. (2000) The 2000 Olympic Games of protein structure prediction; fully automated programs are being evaluated vis-a-vis human teams in the protein structure prediction experiment CAFASP2. *Protein Eng*, **13**(10), 667– 670.
5. Chothia, C. and Lesk, A. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J*, **5**(4), 823–826.
6. Sippl, M. J. and Weitckus, S. (1992) Detection of native-like models for amino acid sequences of unknown three-dimensional structure in a data base of known protein conformations. *Proteins*, **13**(3), 258–271.
7. Gribskov, M., McLachlan, A. D. and Eisenberg, D. (1987) Profile analysis: detection of distantly related proteins. *Proc Natl Acad Sci USA*, **84**(13), 4355–4358.
8. Simons, K. T., Bonneau, R., Ruczinski, I. and Baker, D. (1999) Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins*, **S3,** 171–176.
9. Bayley, M. J., Jones, G., Willett, P. and Williamson, M. P. (1998) GENFOLD: a genetic algorithm for folding protein structures using NMR restraints. *Protein Sci*, **7**(2), 491–499.
10. Eidhammer, I., Jonassen, I. and Taylor, W. R. (2005) *Protein Bioinformatics: An Algorithmic Approach to Sequence and Structure Analysis*. Wiley & Sons, Chichester.
11. Kryshtafovych, A., Milostan, M., Szajkowski, L., Daniluk, D. and Fidelis, K. (2005) CASP6 data processing and automatic evaluation at the protein structure prediction center. *Proteins*, **S7,** 19–23.
12. Cozzetto, D. and Tramontano, A. (2005) Ten years of predictions . . . and counting. *FEBS J*, **272,** 881–882.
13. Hubbard, T. J. (1999) RMS/coverage graphs: a qualitative method for comparing three-dimensional protein structure predictions. *Proteins*, **S3,** 15–21.
14. Moult, J., Fidelis, K., Rost, B., Hubbard, T. and Tramontano, A. (2005) Critical assessment of methods of protein structure prediction (CASP) - round 6. *Proteins*, **S7,** 3–7.
15. Zemla, A. (2003) LGA-a method for finding 3D similarities in protein structures. *Nucleic Acids Res*, **31,** 3370– 3374.
16. Cozzetto, D. and Tramontano, A. (2005) Relationship between multiple sequence alignments and quality of protein comparative models. *Proteins*, **58**(1), 151–157.
17. Valencia, A. (2005) Protein refinement: a new challenge for CASP in its 10th anniversary. *Bioinformatics*, **21**(3), 277–277.
18. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res*, **28**(1), 235–242.
19. Bujnicki, J. M., Elofsson, A., Fischer, D. and Rychlewski, L. (2001) LiveBench-1: continuous benchmarking of protein structure prediction servers. *Protein Sci*, **10**(2), 352–361.

20. Fischer, D., Elofsson, A., Rychlewski, L., Pazos, F., Valencia, A., Rost, B., Ortiz, A. R. and Dunbrack R. L. Jr. (2001) CAFASP2: the second critical assessment of fully automated structure prediction methods. *Proteins*, **S5,** 171–183.
21. Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D. and Tulip, W. R. (1989) Conformations of immunoglobulin hypervariable regions. *Nature*, **342**(6252), 877–883.
22. Tramontano, A., Leplae, R. and Morea, V. (2001) Analysis and assessment of comparative modeling predictions in CASP4. *Proteins*, **S5,** 22–38.
23. DeWeese-Scott, C. and Moult, J. (2004) Molecular modeling of protein function regions. *Proteins*, **55**(4), 942–961.
24. Pizzi, E., Tramontano, A., Tomei, L., La Monica, N., Failla, C., Sardana, M., Wood, T. and De Francesco, R. (1994) Molecular model of the specificity pocket of the hepatitis C virus protease: implications for substrate recognition. *Proc Natl Acad Sci USA*, **91**(3), 888–892.
25. Giorgetti, A., Raimondo, D., Miele, A. E. and Tramontano, A. (2005) Evaluating the usefulness of protein structure models for molecular replacement. *Bioinformatics*, **21**(2), ii72–ii76.