Convergent evolution of protein structure prediction and computer chess tournaments: CASP, Kasparov, and CAFASP

by N. Siew
D. Fischer

Predicting the three-dimensional structure of a protein from its amino acid sequence is one of the most important current problems of modern biology. The CASP (Critical Assessment of Structure Prediction) blind prediction experiments aim to assess the prediction capabilities in the field. A limitation of CASP is that predictions are prepared and filed by humans using programs, and thus, what is being evaluated is the performance of the predicting groups rather than the performance of the programs themselves. To address this limitation, the Critical Assessment of Fully Automated Structure Prediction (CAFASP) experiment was initiated in 1998. In CAFASP, the participants are programs or Internet servers, and what is evaluated are their automatic results without allowing any human intervention. In this paper, we review in brief the current state of protein structure prediction and describe what has been learned from the CAFASP1 experiment, the evolution toward CAFASP2, and how we foresee the future of automated structure prediction. We observe that the histories of "in silico" structure prediction experiments and computer chess tournaments show some striking similarities as well as some differences. We question whether the major advances in automated protein structure prediction stem from novel insights of the protein folding problem, of protein evolution and function, or merely from the technical advances in the ways the evolutionary information available in the biological databases is exploited. We conclude with a speculation about the future, where interesting chess might only be observed in computer games and where the interpretation of the information encoded in the human genome may be achieved mainly through in silico biology.

With the recent advent of the genome sequencing revolution, a flood of information is being accumulated at an exponential rate. The complete genome sequences of a few dozen organisms are available today, dozens more are currently being determined, and within a few years we will have the complete genomes of more than 100 organisms. These range from bacteria and yeast to bigger organisms such as plants, animals, and humans. One of the main aims of collecting all these data is to study the function of the encoded proteins. The eventual goal of all these studies is to understand how the genetic structure affects and controls biological processes. In understanding how diseases and mutations develop, we can perhaps find better ways to deal with them and to design new and better drugs.

However, the mere knowledge of a protein's sequence, or primary structure (1-D), does not allow a detailed understanding of its function. The unique, well-defined, three-dimensional (3-D) structure of a protein dictates the way in which it performs its biological function. Knowing the 3-D structure of a protein allows researchers to gain insight on the active site of the protein or on the way it interacts with small molecules and other proteins. Thus, 3-D structures are essential for a detailed understanding of biology at the molecular level. Although the determination of the complete genome sequences of various organisms has already become routine, the experimental determination of the 3-D structure of the proteins encoded in these genomes is currently a very laborious process. In some cases it can take years before the structure of a protein is determined, and in other

0018-8670/01/$5.00 © 2001 IBM

cases, such as membrane proteins, current methods are not always applicable. In fact, the structure determination step is the bottleneck in the process of fully characterizing a protein. Therefore, only a small fraction of known sequences have a known 3-D structure.

This is where computers come into the picture. With the aid of strong computing power and quantum and statistical mechanics, protein models and simulations of their function could be obtained. The basic assumption is that the information the protein needs in order to fold into its unique 3-D structure lies entirely in its amino acid sequence.[1] It is widely accepted that the native 3-D structure of a protein has the lowest free energy possible for its combination of amino acids. Thus, in principle, finding the unique 3-D structure of a protein given its amino acid sequence alone, is a computable problem. Prediction of protein structure "*in silico*" has thus been the "holy grail" of computational biologists for many years. The aim is to feed the amino acid sequence of a protein to a computer, let it crunch some numbers, and at the end, produce the correct 3-D shape of the protein. However, protein structure prediction *in silico* has proven to be a very difficult task. We do not yet fully understand how a protein folds *in vivo*, nor what are the precise energetic determinants of protein folding.

Interestingly, the histories of *in silico* structure prediction and computer chess show some striking similarities, as well as some differences, and they appear to be an enlightening example of convergent evolution. (Two items are said to be the result of convergent evolution if it is believed that their similarities arose by independent processes without any evidence for common ancestry.) Some common features shared by structure prediction and computer chess are that both are considered "holy grails," both are very complex problems, both require billions and billions of computations, and both are the research topics for hundreds of groups worldwide. The ultimate goal in both fields is to free humans from tedious calculations so that they can concentrate on strategy and tactics and make better use of their expertise and intuition. In both fields it has been questioned whether developments of the automated methods are merely the result of faster, more powerful, algorithms and machines, rather than of major breakthroughs in the understanding of the problems. Another goal shared by both fields has been to apply the developed techniques to other areas of different fields, such as computer science or medicine.

The fields of computer chess and automatic structure prediction also differ in significant ways. Probably one of the major differences lies in the complexity of the rules that govern the games. In chess, the rules are relatively simple and can easily be programmed into a computer. Thus, in principle, given enough computer power, a good level of play can be achieved. In structure prediction the rules are much more complex and are not yet fully understood. Thus, the correct folding of a protein is not ensured even if enough computer power is provided. Other interesting differences relate to the ways in which computers have been used in these fields. Through the years, the field of chess has seen many grand masters playing superior chess without the aid of computers, whereas structure prediction is mainly computer-based; computer chess has reached the level of grand masters, whereas significant improvements in structure prediction methods are required before they achieve a superior performance. In addition, although computer-chess tournaments have been held since the first computer programs appeared, the Critical Assessment of Fully Automated Structure Prediction (CAFASP)[2] experiment is only three years old. In contrast, although three computer-aided structure prediction Critical Assessment of Structure Prediction (CASP)[3] experiments have been held since 1994, computer-aided chess tournaments, where humans equipped with their favorite machines are the participants, are still to be seen.

In this paper we briefly review the field and describe a number of experiments that have been devised to assess progress. Rather than concentrating on computer-chess algorithms or on the parallels of the latter with protein structure prediction algorithms, we focus on the similarities and differences found in the histories of computer-chess tournaments and of protein structure prediction experiments. The paper is thus organized as follows. In the remaining part of the introduction, we briefly define basic protein concepts. In the next section we review current approaches to the protein structure prediction problem and allegorically compare them to computer chess. In the subsequent section we describe the various existing experiments aimed at evaluating the performance of available methods and compare them with computer-chess tournaments. We then conclude with a summary and with some contemplations about the future.

Basic protein concepts. We now briefly introduce some of the protein concepts used in the paper. Proteins are three-dimensional (3-D) molecules that have an important role in all biological processes. The primary structure (1-D) of a protein consists of a chain, or sequence, of amino acids, or residues. Each protein chain folds in space to form the 3-D structure, or fold, of the protein, which in general is uniquely determined by its amino acid sequence. It is the 3-D structure of a protein that dictates the way in which it performs its biological function. The 3-D structure of a protein can inform us of the location of binding sites and of the identity and orientation of active site residues, which can suggest function and reaction mechanisms. This knowledge can aid in rational drug design and protein design. Therefore, knowledge of the 3-D structure of a protein is essential for fully characterizing it and has enormous implications for medicine and human health.

Additional concepts. During the course of evolution a few major processes have occurred. One of them is called "divergent evolution" in which different proteins in different organisms have diverged from a common ancestor protein. Each copy of this ancestor in various organisms has been subject to mutations, deletions, and insertions of amino acids in its sequence, but in general, its 3-D fold and function have remained similar.[4,5] Therefore, two protein sequences that have diverged from the same ancestor can show a certain degree of similarity between them. The similarity between sequences can be observed by aligning them one on top of the other, in such a way that similar regions match, and dissimilar segments are left out as gaps in the alignment. Dynamic programming algorithms are often used to produce the optimal sequence alignment.[6,7] The similarity is often measured by adding the scores of the matching amino acids that occupy the same position in the alignment[8] and by subtracting a penalty for each gap introduced. Percent sequence identity is the percent of identical residues in the alignment. Percent sequence similarity is the percent of similar residues in the alignment. If a long enough alignment has more than 25–30 percent sequence identity and few gaps, it is generally assumed that the two sequences have diverged from the same ancestor, and therefore they are likely to share a similar fold and function.[9,10] If the percent sequence identity is below the 25–30 percent threshold, there are two possibilities. Either the two proteins have diverged from the same ancestor (but their sequences are highly divergent) or the two proteins are unrelated.

Protein structure prediction methods

Computational folding approaches that scan the conformational space, trying to identify those structures with minimal energy, have not yet solved the problem of protein structure prediction. Limitations of such approaches include the vast number of conformations to be scanned, the evaluation of their free energy, and the use of approximations.

Alternative approaches to predict protein structure have recently become a research subfield in bioinformatics.

Alternative approaches to predict protein structure have recently become a research subfield in bioinformatics.[11] These approaches are applicable to special cases and usually employ different principles and various sources of information. These alternative and partial solutions of the protein folding problem are in many cases based on Darwinian and statistical principles. It has become clear that as of today, the theoretical protein folding problem and the more practical protein structure prediction problem are quite different, although much of what can be learned in one field sheds useful insights on the other. The protein folding problem is more commonly studied by theoreticians in the field of physical chemistry, who are mainly concerned with first principles and less concerned about producing working 3-D models for the biologist.[11] In contrast, the protein structure prediction problem is mainly studied by bioinformaticians, a new blend of scientists who have recognized the urgent need of practical solutions for the postgenomic era.

Although modern bioinformatics approaches do not simulate the folding of a protein, they have proven to be efficient and useful. This is an example of a scientific problem that can be (partially) solved in practice, without first obtaining a complete understanding of the protein folding process as it occurs *in vivo*. These partial solutions not only have enormous practical value, but they also entail a significant contribution to the eventual understanding of the protein folding problem. Some of the common approaches used today in order to predict the structure of a protein are *homology modeling*, *fold recognition* (*threading*), and *ab initio*. In what follows we describe each of these approaches very briefly (we refer the reader to a number of recent reviews for further details) and delineate some of the similarities that these approaches share with computer chess.

Homology modeling. Homology modeling, or comparative modeling, is a model-building method based on the Darwinian and empirical principle of "significant sequence similarity implies similarity in 3-D structure."[9,10] Similar protein sequences are assumed to have diverged from a common ancestor. They have accumulated mutations in their sequences, but in general, their function and 3-D structure have been conserved.[4,5] Thus, if the evolutionary relationship between a new target protein and at least one protein of known 3-D structure can be established, a 3-D model for the new protein can be built using as a template the structure of the known protein.[12–14] The process of building a model via homology generally comprises the following six steps.

1. Searching for templates upon which the model could be built. In homology modeling, a template is a protein of known 3-D structure with enough sequence similarity to the sequence of the target protein. If no such template exists, then homology modeling is not applicable, and other approaches need to be used (see below).
2. Aligning the sequences of the target protein and the template protein. The aim of this step is to match each residue in the target sequence to its corresponding residue in the template structure, allowing for insertions and deletions (see e.g., Reference 15).
3. Copying coordinates from the template to the target. With use of the alignment produced in step 2, the coordinates of the matching residues in the known structure are copied, or assigned, to the residues of the unknown protein. In this stage, usually only the backbone coordinates are copied.
4. Building the side chains. Coordinates of identical residues can be imported directly from the template to the target, but the side-chain conformations of the nonidentical ones cannot, and thus they need to be predicted. The prediction of the side-chain coordinates is usually based on empirical data collected from proteins of known structure, such as rotamers,[16,17] and involves solving a difficult combinatoric problem.[18]
5. Building the loop regions. The alignment produced in step 2 includes gaps, usually corresponding to loops between secondary structure elements. Consequently, these loops need to be modeled separately. The missing loop coordinates can sometimes be copied from other known proteins,[19–21] but usually they have to be built *de novo* (e.g., see Reference 22).
6. Improving the model. The model built so far may include steric clashes between atoms because of the different composition of residues in the new protein and the template protein. Therefore, additional adjustments need to be made. Various methods exist for this optimization stage (e.g., see References 23–26).

Homology modeling and computer chess. The best and probably simplest way to draw conceptual similarities between homology modeling and chess might be as follows. The input to a chess program is the current board situation (i.e., a protein sequence). The program searches a library of grand master chess games (a library of proteins with known structure) for a "homologous" board situation that is similar enough to the given input (significant sequence similarity). If such a board situation is found, then the next move that the program chooses is the one recorded in the library. This move is analogous to the stored book of openings in current chess programs. However, it is clear that it is practically impossible to have grand master moves for all possible chess situations, since it is practically impossible to have 3-D structures for every single protein sequence. Thus, if the current board position is absent in the book of openings, the program jumps to the "board-recognition" subroutine (described below).

Despite the fact that we will never have 3-D structures for all proteins, the applicability of homology modeling is enormous. Up to 30 percent of the proteins encoded in the fully sequenced genomes show sufficient sequence similarity to proteins of known structure[27,28] so that good 3-D models can be built for them.

In our allegory above this could mean that the algorithm to find similarities between the input board and the stored boards works well only when the number of differences in the boards (i.e., mutations) is below some threshold. However, it is clear that there are innumerable cases where the number of differences between two chess boards is above our imaginary threshold, although these two boards still represent a very similar game (i.e., 3-D structure).

Consequently, this implies that the algorithm that detects board similarities by simply counting the number of differences between two boards has not yet captured the full essence of the chess game, and a more sensitive algorithm that detects distant board relationships is needed.

Fold recognition. In the lack of significant sequence similarity, a new target protein may still be structurally similar to one of the proteins of known 3-D structure.[29,30] It has been estimated that over 50 percent of genome proteins will have 3-D structures similar to protein folds already observed.[31] Because it is not possible to identify the correct template for the majority of these cases with standard sequence comparison techniques, more sensitive methods are needed. One of these methods is known as fold recognition, or threading. Fold recognition is aimed at identifying a correct template structure for those prediction targets that show no significant sequence similarity to any of the proteins of known structure. If such a template exists, a further goal of fold recognition is to provide an accurate sequence-structure alignment between target and template. That is, fold recognition replaces homology modeling steps 1 and 2, described previously. The approach used by fold recognition is to measure *sequence-structure compatibility* rather than mere sequence similarity as in homology modeling.[32] Fold recognition methods vary mainly in the way in which the sequence-structure compatibility is measured, but they generally share five essential components:

1. A library of known three-dimensional folds[33–35]
2. A representation of the 3-D information of the library folds in a way suitable for the sequence-to-structure compatibility function (see References 32, 36, and 37)
3. A sequence-to-structure compatibility function that scores the compatibility of a sequence to a given fold. This function is a mapping of the one-dimensional sequence of the target onto the 3-D information of a protein fold. The compatibility function can take into account a number of features that include the preferences of the amino acids to be in different structural environments[36–41] (for recent reviews see, e.g., References 32, 42, and 43).
4. A method to optimally "thread," or align, the target sequence into the 3-D fold, using the sequence-to-structure compatibility function. Computing the optimal gapped alignment is an NP (nondeterministic polynomial time) -complete problem if the compatibility function takes into consideration pair interactions.[32,44] Thus, in these cases, approximations and heuristics need to be used.[39,45,46]
5. A method to assess significance. Each alignment of the target with a fold library receives a compatibility score. The magnitude of the scores depends, among other things, on the lengths and composition of the target and fold sequences and on the number of folds in the library. To select the best template candidate, an assessment of the significance of the top scores is required.[32,47]

Fold recognition and computer chess. In our allegorical analogy to chess, fold recognition corresponds to a hypothetical algorithm, which could be termed "board recognition." This algorithm would look for essential similarities between a given board situation and the stored boards in a library, which go well beyond simple variations. This algorithm needs to be more sensitive than the homology modeling one, which only looks for almost identical arrangements of the chessboard. If the more sensitive algorithm detects a "known" board that can serve as a template, then the next move is deduced from the stored move. If not, then the chess program would jump to the *ab initio* subroutine described below. Whether such an algorithm may exist for chess is beyond the scope of this paper and the expertise of the authors. If an algorithm to detect the essence of chess situations existed, and we knew that the number of "representative" chess situations is finite and relatively small, then, in principle, it would be trivial to produce a program that plays outstanding chess. Although in chess this may not be feasible because the number of possible situations is too large, in protein structure prediction we have a brighter situation. It is believed that the total number of protein folds in nature is finite and relatively small—between one to a few thousand folds[29,48]out of which around 600 to 700 are currently known[49,50]—and this number is growing fast. However, the number of sequences without significant sequence similarity to proteins of known structure is orders of magnitude larger. Soon, most of the prediction targets will probably be identified as having one of the already known folds. Consequently, fold recognition methods have an enormous applicability, which will grow as we approach the "1000" folds mark. The same may be true to some extent for homology modeling. However, in order to have a 3-D structure within "homology-modeling-distance" for the majority of genome proteins, the 3-D structures of hundreds of thousands of proteins may be required, because throughout the course of evolution sequences have diverged faster than structures. Structural genomics[51–56] aims to fill the gaps in our structural knowledge by attempting to determine the structures of each sequence family, and to prioritize efforts to determine structures with likely novel, previously unobserved folds.[27,55] Thus, until one representative structure exists for each sequence family, fold recognition is the method of choice when a target protein has no significant sequence similarity to a known structure.

***Ab initio*.** Since our structural knowledge does not yet approach a complete mapping of all the folds in nature, we will continue to observe a non-negligible number of genome sequences that cannot be modeled using the currently known structures, and thus neither homology modeling nor fold recognition can be used to predict their structures. For this purpose, a number of methods that do not directly rely on known 3-D structures have been developed, and they usually are referred to as "*ab initio*"[57] methods.[43,58–62]

In general, *ab initio* methods are composed of the following three essential components (see Reference 63 and references within):

1. A representation of protein geometry. Because all-atom models of the protein and the solvent environment are computationally expensive, a number of approximations in representing the protein and the solvent are used. These include methods using one or a few atoms per residue and an implicit solvent, e.g., see References 64–67.
2. A potential energy function and other parameters that are generally based on statistical analysis of known structures of proteins;[68,69] for a review see, e.g., Reference 70.
3. A conformational search technique. The majority of current *ab initio* methods search the energy surface using methods such as Monte Carlo, simulated annealing, genetic algorithms, or molecular dynamics.[24,25,66,71–73]

Ab initio and computer chess. The *ab initio* problem is difficult because in each of its three components large approximations are required. In addition to requiring enormous computing power, these methods also suffer from inaccuracies inherent in current potential functions.[32,63] Thus, an enormous amount of research is currently being carried out to improve both the search techniques and the potentials used.[67,74–76]

In computer chess, if enough computing power were available, a chess program could compute all possible moves until the end of the game and always choose the best one. In this case, the evaluation function is simply to assess whether a checkmate was reached or not. In this sense, and unlike protein structure prediction, it is obvious that unlimited computer power alone would solve the problem. However, because it is impossible to compute such a large number of moves, modern chess programs "only" look ahead a few dozen moves and apply sophisticated evaluation functions that assess the value of the positions reached. These functions are not perfect, of course, and neither are the potential functions used in *ab initio* methods. Another important similarity between computer chess and *ab initio* protein structure prediction methods is that in both fields intensive algorithmic development has been carried out in order to achieve a faster, more complete sampling of the search space.

Evaluating prediction methods

After a model is produced with any of the structure prediction methods, how do we know how similar to the real structure the predicted model is? How do we measure the similarities between the model and the real structure of the protein? In other words, how do we know how "good" the current structure prediction methods are? Addressing these questions has already become an intense subfield of research.

To test the performance of their methods, researchers need a way to compare the predicted models to the real structures. However, the real structures of prediction targets are, by definition, unavailable. Thus, researchers usually test their methods by applying them to known structures, pretending that they are prediction targets of unknown structure, and then checking how similar the prediction is to the real structure. This "postdiction" testing is not a blind prediction, because it is possible that biases are introduced, consciously or unconsciously.

In this section we describe three ongoing, worldwide experiments that try to deal with the evaluation of prediction methods and discuss some of the lessons that were learned from them.

CASP. The idea to test the prediction methods in a blind manner, which enables a direct comparison of a protein model to its real structure, was the basis for the CASP experiments, initiated in 1994 by John Moult, of the Center for Advanced Research in Biotechnology located in Maryland. In CASP, a few dozen proteins of known sequence but unknown structure are used as prediction targets. Contestants are asked to file their predictions before the real 3-D structure of the protein is experimentally determined. The predictions are filed using various methods spanning the three main categories: homology modeling, fold recognition, and *ab initio*. Subsequently, when the 3-D structure is released, an assessment of the accuracy of the predictions is carried out. This protocol ensures that no participant knows the correct answer while building a model and, thus, the submitted responses effectively reflect a blind prediction at the time of the contest. The models in CASP are usually produced by a combination of computer programs and human intervention.

Researchers need a way to compare the predicted models to the real structures.

CASP is held every two years and concludes with a meeting in Asilomar, California, to discuss the results.[3,77,78] The CASP4 meeting was held in December 2000, and over 150 predicting groups worldwide participated. Over the course of the first three CASP events held so far (at the time this paper was written, June 2000, CASP4 was taking place), much was learned about the strengths and weaknesses of the various approaches and methods of structure prediction. In the homology modeling section, often very good models are produced, providing "detailed hypotheses of catalysis, ligand binding, and allosteric regulation" and suggesting "which residues to mutate in experimental tests."[3] It was also shown that a key factor to achieving a good model is to have a good alignment between the unknown protein and the known proteins.[79,80] This step is most crucial to the building of a model. When the sequence identity between the target protein and the template structure is above 70 percent, current methods seem to have no problem in finding the correct alignment. However, in the cases where the similarity between the proteins is around 20–25 percent, it is very hard to achieve a correct alignment. Such a low level of similarity is sometimes referred to as the "twilight zone" of sequence similarity.

When there is less than 20 percent identity between the target and the best template, homology modeling methods perform very poorly, mainly because it is hard to find correct template structures. Even if such a template is found, it is likely that the alignment will contain large errors.[81] Although the challenge is to produce models in which there is an advancement in the accuracy and detail beyond the mere copying of the template coordinates,[3] the accuracy of models depended on the amount of information directly transferred from the template to the target.[81] Another clear finding from CASP was that current optimization methods usually do not improve the models and can even make them worse.[81] Thus, there is still room for improvement until we reach a level of accuracy that rivals experimental structures.[3]

As for the fold recognition category, the CASP experiments have clearly demonstrated that threading methods succeed for those targets for which there is a known structure belonging to the same protein family (divergent sequences with a probably common ancestor). But for those proteins for which the closest template belongs to a different family (convergent structures with probably no common ancestor), it is not yet clear how well fold-recognition methods perform.[3,43,82] Other conclusions from CASP were that the quality of alignments for the medium difficulty targets has increased[3] and that many fold recognition methods have become hybrid methods.[43,80,83,84] These hybrid methods combine structure information with various types of sequence information from multiple alignments,[85,86] and also with the use of predicted secondary structure (e.g., Reference 58).

In the *ab initio* category, the CASP experiments have shown that useful models cannot yet be produced, although the methods are constantly advancing. In CASP1 and CASP2 there were almost no good *ab initio* predictions, but in CASP3 a few groups produced excellent models;[68,69,87–90] for recent reviews see, e.g., References 91 and 92.

Measuring progress through the previous three CASPs is very important in order to learn about the developments and advancements of the structure prediction methods. However, progress in CASP is difficult to assess for a number of reasons. First, the number of targets is relatively small and, therefore, the results may not always be significant. Second, the assessment is carried out by humans, with different assessors in each CASP and, therefore, it is not always straightforward to reproduce. Third, the difficulty of the targets used in the evaluation was different in each CASP. Thus, it is clear that in order to assess progress over the years, large-scale experiments with homogeneous, reproducible, and automated evaluation methods are needed. Nevertheless, from various analyses[3,93–95] it appears that, in general, significant improvements can be observed from CASP1 to CASP2, but in CASP3 the improvement appeared to be less dramatic.

CAFASP. Despite the enormous value of the CASP experiments, they do have some limitations, one of which is that CASP can only assess the performance of computer-*aided* structure prediction. Since human intervention is allowed when producing the predictions, what is measured are the capabilities of human experts using prediction programs and not the capabilities of the programs themselves. However, assessing the performance of fully automatic methods is critical for biologists. When biologists aim to predict the structure of a protein, what they wish to know is which program performs best and not which group was able to produce the best predictions at CASP. With the advent of genome sequencing projects, including the human genome, the need for fully automated structure prediction has become evident. A few years ago, automated tools were either nonexistent or highly inaccurate. But as protein structure prediction has evolved and a number of automated tools have demonstrated that they are already able to produce valuable predictions in many cases, it became important to test their capabilities alone.

The benefits of an assessment of fully automated methods are manifold. First, the nonspecialist users can choose which is the best method for them to use on their prediction targets. Second, users can evaluate and better interpret the results they obtain from the various prediction programs. And last, fully automated predictions are reproducible, unlike the cases where human intervention is part of the model-building process.

To address these needs, the CAFASP experiment was initiated by Daniel Fischer from our group (see http://www.cs.bgu.ac.il/~dfischer/CAFASP2). CAFASP1 was a small experiment with only a handful of fold-recognition servers. The prediction targets were the same as those used in CASP3, but the experiment took place after the real structures were revealed, so that in a sense the prediction was not fully blind. However, the models were produced completely automatically without any human intervention. The CAFASP1 results demonstrated that although in most cases human intervention resulted in better predictions, several programs could already independently produce reasonable models.

CAFASP2 was run in parallel to CASP4 using the same prediction targets. The target sequences were sent automatically to the participating servers at the time they were released for the CASP4 participants, and the data produced by these servers are automatically collected and stored for subsequent evaluation. The developers of the servers were not active in this process, and therefore an assessment of fully automated, blind predictions could be achieved. At the time this paper was initially written (June 2000) CAFASP2 was taking place with over two dozen registered automatic servers from five continents (see Table 1). CAFASP2 covered all aspects and methods of automated protein structure prediction, including the one considered to be the most difficult: *ab initio*. The first fully automated *ab initio* servers were two of the CAFASP2 participants. Figure 1 is an example from CAFASP2. Members of the prediction community, and in particular the nonexpert protein structure predictors in the wider biology community, are waiting to learn about the capabilities of automated structure prediction. By the time this paper is published, the results of the experiment will be available at the CAFASP Web site listed above.

Protein structure experiments and computer chess tournaments. The CAFASP experiment resembles the computer chess tournaments that have existed since computer chess began. It was obvious to chess program developers that in order to learn which are the best existing programs, a tournament was needed. It was also clear that the programs should play alone, without the expert input from humans; otherwise it would not be possible to distinguish whether the success or failure of a program was due to the program itself or due to the human that intervened. As for protein structure prediction, the state of the field up until recently was such that a fully automated protein structure prediction tournament would have shown that the programs were not very successful, to say the least. But, as prediction programs have improved, the time became ripe to begin such tournaments. We hope that the experience gained through years of computer chess tournaments will be useful in our future CAFASP experiments.

Because the CAFASP experiments may gain valuable insights from the experience obtained in computer chess tournaments, it may also be that the computer chess community will benefit from the experience gained in the CASP experiments, where computer-

Table 1 Protein structure prediction servers registered at CAFASP2

Server Name	Type*	URL
		United States
PHD/PROF (Rost)	SS/SS	http://dodo.cpmc.columbia.edu/predictprotein
SSpro	SS	http://promoter.ics.uci.edu/BRNN-PRED
SDSC1	HM	http://c1.sdsc.edu/hm.html
FFAS	FR	http://bioinformatics.burnham-inst.org/FFAS
SAM-T99	FR/SS	http://www.cse.ucsc.edu/research/compbio/MMM-apps/T99-query.html
P-Map	FR	http://www.dnamining.com
loopp	FR	http://ser-loopp.tc.cornell.edu/loopp.html
123D+	FR	http://www-lmmb.ncifcrf.gov/~nicka/123D+.html
Isites	AI	http://honduras.bio.rpi.edu/~isites/ISL_rosetta.html
Dill-Ken	AI	http://www.dillgroup.ucsf.edu/~kdb
		Asia
PSSP	SS	http://imtech.ernet.in/raghava/pssp
FAMS	HM	http://physchem.pharm.kitasato-u.ac.jp/FAMS
rpfold	FR	http://imtech.chd.nic.in/raghava/rpfold
		Europe
Jpred2	SS	http://jura.ebi.ac.uk:8888
Pred2ary	SS	http://www.cmpharm.ucsf.edu/~jmc/pred2ary/
PROF (King)	SS	http://www.aber.ac.uk/~phiwww/prof/index.html
Nanoworld	SS	http://ftp.decsy.ru/nanoworld/DATA/PROGRAMS/DNETOENT/dne_to_ent.htm
3D-JIGSAW	HM	http://www.bmm.icnet.uk/people/paulb/3dj
GenTHREADER/Psipred	FR/SS	http://www.psipred.net
3D-PSSM	FR	http://www.bmm.icnet.uk/servers/3dpssm
FUGUE	FR	http://www-cryst.bioc.cam.ac.uk/~fugue
ssPsi	FR	http://130.237.85.8/~arne/sspsi
threadwithseq	FR	http://montblanc.cnb.uam.es
CORNET	CP	http://prion.biocomp.unibo.it/cornet.html
PDG_contact_pred	CP	http://montblanc.cnb.uam.es/cnb_pred/pdg_contact_pred.html
		Middle East
bioinbgu	FR	http://www.cs.bgu.ac.il/~bioinbgu/
		Australia
Sausage	FR	http://rcs.anu.edu.au/~arussell/TheSausageMachine.html

*Type of server: SS stands for Secondary Structure Prediction
HM stands for Homology Modeling
FR stands for Fold Recognition
AI stands for *ab initio*
CP stands for contacts predictions

For further details see http://cafasp.bioinfo.pl/server/

aided structure prediction is assessed. Personally, we would be anxiously watching the first computer-aided chess tournament, where human grand masters equipped with their favorite programs play against each other. And in particular, one eagerly awaited event might be a tournament where a program plays against a human equipped with a machine. Such events may provide invaluable insights to understanding what humans add to the game that machines do not use in their calculations, and whether this human expertise is computable at all and thus amenable to being incorporated into programs. This aspect is probably one of the most interesting of artificial intelligence. The analysis of such a comparison will be one of the most important outcomes of the CASP4 and CAFASP2 experiments.

Questions raised by the emergence of the protein structure experiments. Here we discuss four of the questions that have been raised during the past protein structure experiments that have important implications within and beyond the protein structure prediction field.

1. How do we compare human versus machine performance? One of the most anxiously awaited results of CASP4 and CAFASP2 is the comparative analysis of the performance of humans (CASP4) with that of the automatic programs (CAFASP2). The performance comparison of humans versus machines will allow the amount of human intervention required in current interactive predictions to be objectively quantified for the first time. Understanding and analyzing the aspects of human expertise that lead to a better human performance will allow their future incorporation into automated programs; this challenge is and will continue to be a major one for developers. In the CAFASP2 and CASP4 comparison it is expected that humans will perform better, in part because the automated predictions from the servers are available long before the filing deadline for the human predictions, and human predictors can make use of the automated results when preparing their predictions (but not vice versa).

Comparing human and machine performance is beginning to raise interest similar to that of the man-versus-computer matches in chess. It took over 20 years of computer chess tournaments before a machine beat a grand master. Although machines will probably not outperform humans this year, we should not bet high against machines in CAFASP3.

2. Is automation a valid scientific enterprise? Some objections have been raised as to the scientific value of the recent developments in automation. In both fields it has been questioned whether developments of the automated methods are merely the result of faster, more powerful algorithms and machines, and of larger databases, rather than of major breakthroughs in understanding the problems with which we are dealing. From the theoretical point of view, no major breakthroughs in understanding have been achieved from efforts to improve the techniques. Current programs are far from being able to simulate the folding of a protein as it occurs *in vivo*, or to simulate the thought processes that occur within the brain of a human chess grand master. The development of automated tools appears to have led to only a slightly better understanding of the above problems. In particular, it was clear from the CASP experiments that very little effort was put toward solving the protein folding and stabilization problems, which require addressing the physics-based aspect of protein structure. Perhaps the reason is because computing these important scientific aspects is still very complicated.[3] Moult et al. have also accurately observed that current methods tested at CASP mainly focus on technology development for producing the best possible protein structure models, rather than on basic science.[3] A similar process has been observed in other fields; long periods of technology development were required before major advances in understanding were achieved.[3] We thus expect that after further developments in protein structure technology are achieved, we will begin to understand the basic aspects of the protein folding process in a more profound way.

3. Why automate structure prediction? Considerable benefits are to be gained from studies of automated structure prediction. Given the large amount of raw data available today, the need to obtain as much information on unknown proteins as possible, and the improvements in automated structure prediction, automation will play a major role, especially at the initial, genomic-scale, screening steps. The goal, as in chess, is to encourage further development of the automated tools so that they become more routine companions in the prediction tasks, ridding humans from as many tedious computations as possible and allowing them to better apply their intuition and expertise. If something is computable, programs should be written to compute it, and their performance should be thoroughly tested. With good automatic procedures, human experts will be free to concentrate on the more important questions and aspects of protein research, such as the fold pathway, the forces that determine the fold, and, of course, other biological and pharmaceutical aspects. In addition, improvements in automated structure prediction will allow us to distinguish more and more cases of accurate and reliable predictions. This will leave fewer cases for human intervention—a most important goal in the postgenomic era.

4. Will machines replace humans? The fact that a machine begins to compete with humans, and may eventually beat them, should be regarded as a great accomplishment for humanity. Rather than a loss, it is a celebration of a human's capabilities; humans created the machine, after all. However, in the near future, it is likely that human intervention will still be required to improve the automatically generated 3-D protein structure models, because of the knowledge, expertise, and intuition that humans have and that programs still lack. Furthermore, we are still far away from the time when computer 3-D models will

be able to replace experimentally determined protein structures.

In chess, perhaps the situation is a bit different. It is likely that in the future, humans (without the aid of computers) will not be able to beat a machine. Nevertheless, this does not mean that the interest of human chess will disappear, and not only for recreational purposes. The character of the game of a human grand master will probably continue to be for the near future far more interesting than that performed by computers. No superiority of computer over human has been achieved as far as technique and style is concerned. Since in chess the strategy and "elegance" of thinking are essential parts of the game, computer chess algorithms still need to be improved in these directions. We should perhaps wait to see chess tournaments that evaluate style and technique in addition to plain victory.

LiveBench. A limitation of both CASP and CAFASP is the relatively small number of prediction targets (a few dozen). To overcome this limitation, a large-scale evaluation of automatic servers, named "LiveBench," was recently initiated by Leszek Rychlewski from Poland. LiveBench follows the CAFASP ideology in that it evaluates automatic servers only, and it works as follows. Each week the Protein Data Bank (PDB)[96] is checked for new entries. Proteins with low sequence similarity to other proteins of known structure are chosen as prediction targets for LiveBench and are immediately submitted via the Internet to the participating servers. After a few months, a large collection of prediction targets is thus obtained, and the predicted models can be evaluated. Although LiveBench uses new PDB releases, it practically entails a blind experiment, because it is highly unlikely that developers "adjust" their servers on a weekly basis only to improve their performance at LiveBench.

LiveBench-1 is currently under way, with only a handful of fold-recognition servers. Preliminary results show that the best servers are able to produce correct models for between one-third and one-half of all newly released structures that show no sequence similarity to other proteins of known structure[97] (see http://bioinfo.pl/LiveBench for further details). These results show that automated servers have a significantly higher sensitivity than standard sequence-based methods. Unfortunately, this increase in sensitivity came with the cost of lower specificity. Another interesting finding of LiveBench was that an "ideally combined consensus" of all servers would increase the percentage of correct assignments by 50 percent. This hints at the benefits of using more than one server for difficult prediction targets. A similar, large-scale evaluation of secondary structure prediction methods is led by Burkhard Rost from Columbia University (see http://maple.bioc.columbia.edu/eva for details).

The main contribution of large-scale evaluation experiments is, like CAFASP, to inform biologists about the current performance of available automated servers; the main difference between them and CAFASP is that the large-scale projects are carried out in a continuous fashion and use a larger number of prediction targets.

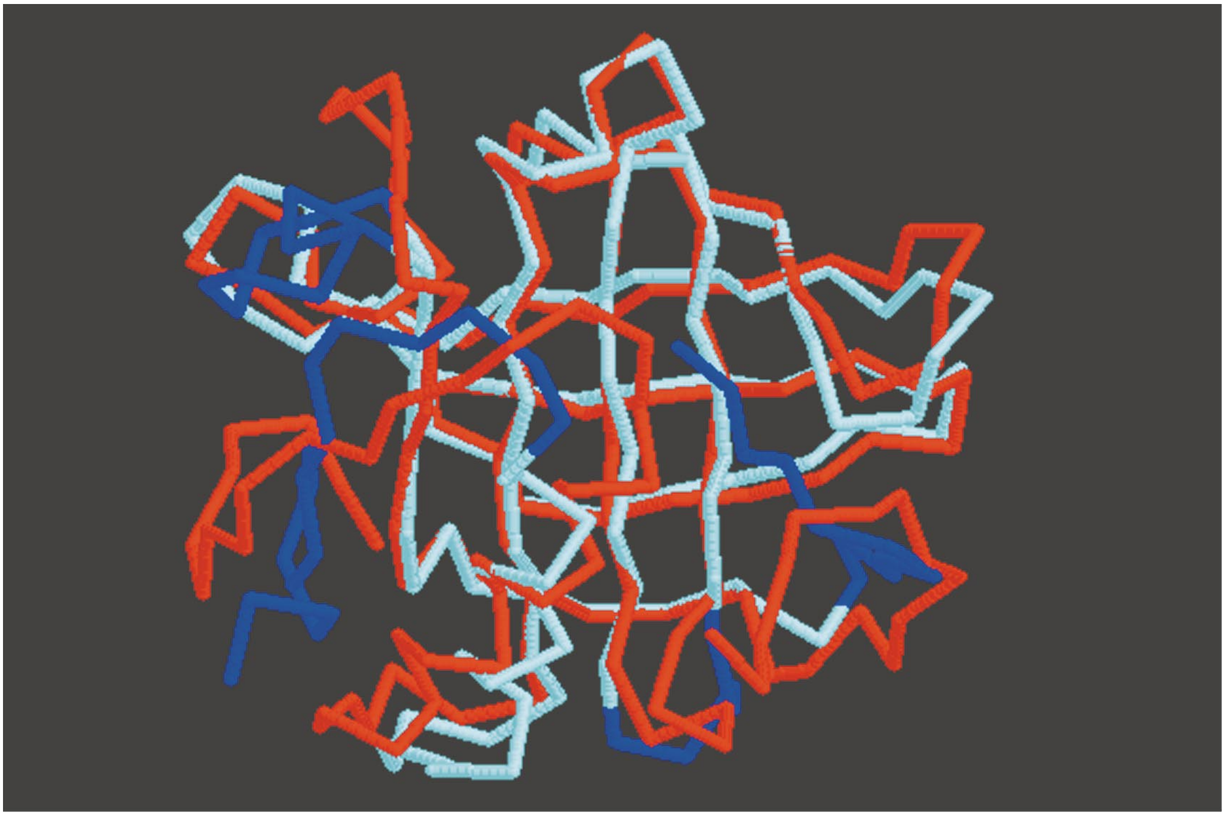
The problem of model evaluation. During the CASP experiments it became clear that the evaluation process of the predicted 3-D models vis-à-vis the real structures is a difficult problem. In CASP, different criteria were used for the assessment, partly automatic, partly involving human expertise and knowledge. Each criterion focused on different aspects of a 3-D model. Evaluating how good a predicted 3-D model is has turned out to be a controversial subfield of research.

In CAFASP, an objective, fully automated, quantitative, and reproducible evaluation method is used. To this end, a single numerical measure that can be added over all predictions was needed so that an estimation of the overall performance of each method can be obtained. To address these needs, the MaxSub method was developed.[98] MaxSub measures the quality of a predicted model by searching for the largest subset of C-alpha atoms in the model that superimpose well over the real structure of the protein (see Figure 1), and by producing a normalized score that reflects the quality of this superimposition. Although this problem is difficult, we have shown that a heuristic approach performs well.[98]

The availability of such an automated evaluation measure allows large-scale evaluation experiments, such as LiveBench, to take place. It also allows the achievement of full automation in CAFASP2 and in LiveBench, both in the way the models are produced and in the way they are evaluated.

Evaluation experiments such as CASP, CAFASP, and LiveBench are becoming the standard measures of progress in the field and effectively reflect the state of the art of structure prediction at the time of the experiment. The value of such experiments is enor-

Figure 1 Superposition of a predicted protein model from CAFASP2 (in cyan and blue) with the experimental protein structure (in red)



The model was automatically produced by one of the CAFASP2 servers. The superposition was computed by MaxSub,[98] which identified 113 "well-predicted" residues (cyan) that superimpose onto the experimental structure with a root-mean-square deviation of 3.7 Å. Portions of the model that were not well-predicted are in blue (a total of 43 residues). This prediction has a MaxSub score (using a scale of zero to one, where zero corresponds to a wrong prediction and one corresponds to a perfect prediction) of 0.6, which corresponds to a relatively good prediction, where most of the secondary structures match well. In the absence of an experimental structure, this level of accuracy in a predicted model can often be very helpful for biologists.

mous: They show the strengths and weaknesses of each method and encourage developers to improve their programs. They also inform researchers outside the prediction community, including biologists and commercial companies, about the capabilities, limitations, and progress of current structure prediction methods. And finally, they have catalyzed significant improvements in automated structure prediction so that current methods have already become routine companions of many biologists.

Summary and discussion

Progress in automation has definitely changed both the computer chess and protein structure prediction fields. With the availability of automation, both fields have become more interesting as deep and challenging questions, impossible to address before, can now be explored. Computer programs in this and other fields, rather than challenging humans, have made a better world. From a practical point of view, automated protein structure prediction is changing the methodology of biology. Current tools already allow for many useful predictions and are becoming routine tools for biologists (for a few recent examples see http://www.cs.bgu.ac.il/~bioinbgu). Further advances in the prediction methods, especially homology modeling and fold recognition, will become increasingly important and widely applicable as the various structural genomics projects continue to fill the gaps in our structural knowledge.[51–56] The importance of being able to predict the structure of the proteins encoded in the genomes of various organisms is enormous. The means to achieve this prediction via technological or basic science advances will ultimately be of minor importance. If there is room for technological improvement in the near future so that available information is collected and more accurate models are produced, it is essential that bioinformaticians fulfill this need. Their contribution will be to the biological sciences first, and only later to the biophysical aspects of protein folding.

The CAFASP and LiveBench experiments are a contribution toward the long-sought-after goal of being able to submit to a computer the complete genome sequence of an organism, and upon a number of calculations, obtain the 3-D structures of each of the encoded proteins. Comparing the performance of humans versus that of automated programs will help us learn what it is that humans know and machines do not. The result will be a major step toward our ability to understand the relationship between structure and function in biological systems, to prevent and cure diseases, and to control processes in living systems. Although these goals will not be fully achieved in CAFASP2, subsequent tests will serve as further catalysts to this process and as measures of continuing success. The protein structure prediction community and the wider community of bioinformaticians and biologists using these tools will certainly be watching the 2000 protein structure prediction Olympic games[99] for the advances in the classic "human-plus-machine" CASP category, for the new reports of the fully-automated CAFASP category, and for the comparison between the two.

Automation efforts in one field can sometimes lead to unexpected advances in other, nonrelated fields. One such example is the development of the Blue Gene project,[100,101] which will probably make use of many lessons learned in the Deep Blue project. Similarly, it may not be completely absurd to imagine a program that, having evolved from principles used in the protein structure prediction field, will become the future world chess champion.

For more information see the CASP site at http://PredictionCenter.llnl.gov/casp4, the CAFASP site at http://www.cs.bgu.ac.il/~dfischer/CAFASP2, and the LiveBench site at http://bioinfo.pl/LiveBench.[102,103]

Acknowledgment

Naomi Siew is partly supported by the Kreitmann Foundation and the Council for Higher Education High Technology fellowships.

Cited references and notes

1. C. Anfinsen, "Principles That Govern the Folding of Protein Chains," *Science* **181**, 223–227 (1973).
2. D. Fischer, C. Barret, K. Bryson, A. Elofsson, A. Godzik, D. Jones, K. J. Karplus, L. A. Kelley, R. M. MacCallum, K. Pawlowski, B. Rost, L. Rychlewski, and M. Sternberg, "CAFASP-1: Critical Assessment of Fully Automated Structure Prediction Methods," *Proteins: Structure, Function, and Genetics* Supplement **3**, 209–217 (1999).
3. J. Moult, T. Hubbard, K. Fidelis, and J. T. Pedersen, "Critical Assessment of Methods of Protein Structure Prediction (CASP): Round III," *Proteins: Structure, Function, and Genetics* Supplement **3**, 2–6 (1999).
4. C. Chothia and A. M. Lesk, "Relationship Between the Divergence of Sequence and Structure in Proteins," *EMBO Journal* **5**, 823–827 (1986).
5. A. M. Lesk and C. Chothia, "The Response of Protein Structure to Amino-Acid Sequence Changes," *Philosophical Transactions of the Royal Society of London* **317**, 345–356 (1986).
6. S. B. Needleman and C. D. Wunsch, "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins," *Journal of Molecular Biology* **48**, 443–453 (1970).
7. T. F. Smith and M. S. Waterman, "Identification of Common Molecular Subsequences," *Journal of Molecular Biology* **147**, 195–197 (1981).
8. M. O. Dayhoff, W. C. Barker, and L. T. Hunt, "Establishing Homologies in Protein Sequences," *Methods in Enzymology* **91**, 524–545 (1983).
9. C. Sander and R. Schneider, "Database of Homology-Derived Protein Structures and the Structural Meaning of Sequence Alignment," *Proteins: Structure, Function, and Genetics* **9**, 56–68 (1991).
10. M. Hilbert, G. Bohm, and R. Jaenicke, "Structural Relationships of Homologous Proteins as a Fundamental Principle in Homology Modeling," *Proteins: Structure, Function, and Genetics* **17**, 138–151 (1993).
11. B. Honig, "Protein Folding: From the Levinthal Paradox to Structure Prediction," *Journal of Molecular Biology* **293**, 283–293 (1999).
12. T. A. Jones and S. Thirup, "Using Known Substructures in Protein Model Building and Crystallography," *EMBO Journal* **5**, No. 4, 819–822 (1986).
13. A. Sali, "Modeling Mutations and Homologous Proteins," *Current Opinion in Biotechnology* **6**, No. 4, 437–451 (1995).
14. M. S. Johnson, N. Srinivasan, R. Sowshamini, and T. L. Blundell, "Knowledge-Based Protein Modeling," *CRC Critical Reviews in Biochemistry and Molecular Biology* **29**, 1–68 (1994).
15. G. J. Barton, "Protein Sequence Alignment and Database Scanning," *Protein Structure Prediction: A Practical Approach*, M. J. E. Sternberg, Editor, IRL Press at Oxford University Press, Oxford (1996), pp. 31–64.
16. J. Janin, S. Wodak, M. Levitt, and B. Maigret, "Conformation of Amino Acid Side Chains in Proteins," *Journal of Molecular Biology* **125**, 357–386 (1978).
17. J. W. Ponder and F. M. Richards, "Tertiary Templates for Proteins: Use of Packing Criteria in the Enumeration of Allowed Sequences for Different Structural Classes," *Journal of Molecular Biology* **193**, 775–791 (1987).
18. M. Vasquez, "Modeling Side-Chain Conformation," *Current Opinion in Structural Biology* **6**, No. 2, 217–221 (1996).
19. J. Moult and M. N. G. James, "An Algorithm for Deter-

mining the Conformation of Polypeptide Segments in Proteins by Systematic Search," *Proteins: Structure, Function, and Genetics* **1**, 146–163 (1986).

20. R. E. Bruccoleri and M. Karplus, "Prediction of the Folding of Short Polypeptide Segments by Uniform Conformational Sampling," *Biopolymers* **26**, 137–168 (1987).
21. K. Fidelis, P. S. Stern, D. Bacon, and J. Moult, "Comparison of Systematic Search and Database Methods for Constructing Segments of Protein Structure," *Protein Engineering* **7**, 953–960 (1994).
22. V. Collura, J. Higo, and J. Garnier, "Modeling of Protein Loops by Simulated Annealing," *Protein Science* **2**, 1502–1510 (1993).
23. N. Srinivasan, K. Guruprasad, and T. L. Blundell, "Comparative Modelling of Proteins," *Protein Structure Prediction: A Practical Approach*, M. J. E. Sternberg, Editor, IRL Press at Oxford University Press, Oxford (1996), pp. 111–140.
24. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, B. J. States, S. Swaminathan, and M. Kaplus, "CHARMM: A Program for Macromolecular Energy Minimization and Dynamics Calculations," *Journal of Computational Chemistry* **4**, 187–217 (1983).
25. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science* **220**, 671–680 (1983).
26. L. Holm and C. Sander, "Fast and Simple Monte Carlo Algorithm for Side Chain Optimization in Proteins: Application to Model Building by Homology," *Proteins: Structure, Function, and Genetics* **14**, 213–223 (1992).
27. D. Fischer and D. Eisenberg, "Predicting Structures for Genome Proteins," *Current Opinion in Structural Biology* **9**, 208–211 (1999).
28. R. Sanchez and A. Sali, "Large-Scale Protein Structure Modeling of the *Saccharomyces cerevisiae* Genome," *Proceedings of the National Academy of Sciences (USA)* **95**, 13597–13602 (1998).
29. C. A. Orengo, D. T. Jones, and J. M. Thornton, "Protein Superfamilies and Domain Superfolds," *Nature* **372**, 631–634 (1994).
30. C. A. Orengo, T. P. Flores, D. T. Jones, W. R. Taylor, and J. M. Thornton, "Recurring Structural Motifs in Proteins with Different Functions," *Current Biology* **6**, 131–139 (1993).
31. D. Fischer and D. Eisenberg, "Assigning Folds to the Proteins Encoded by the Genome of *Mycoplasma genitalium*," *Proceedings of the National Academy of Sciences (USA)* **94**, 11929–11934 (1997).
32. D. Fischer, D. Rice, J. U. Bowie, and D. Eisenberg, "Assigning Amino Acid Sequences to 3-Dimensional Protein Folds," *FASEB Journal* **10**, 126–136 (1996).
33. U. Hobohm, M. Scharf, R. Schneider, and C. Sander, "Selection of Representative Protein Data Sets," *Protein Science* **1**, 409–417 (1992).
34. J. Boberg, T. Salakoski, and M. Vihinen, "Selection of a Representative Set of Structures from Brookhaven Protein Databank," *Proteins: Structure, Function, and Genetics* **14**, 265–276 (1992).
35. D. Fischer, C. J. Tsai, R. Nussinov, and H. Wolfson, "A 3-D Sequence-Independent Representation of the Protein Data Bank," *Protein Engineering* **8**, No. 10, 981–997 (1994).
36. J. U. Bowie, R. Luthy, and D. Eisenberg, "A Method to Identify Protein Sequences that Fold into a Known Three-Dimensional Structure," *Science* **253**, 164–170 (1991).
37. M. J. Sippl, "Calculation of Conformational Ensembles from Potentials of Mean Force: An Approach to the Knowledge-Based Prediction of Local Structures in Globular Proteins," *Journal of Molecular Biology* **213**, 859–883 (1990).
38. M. J. Sippl and S. Weitckus, "Detection of Native Like Models for Amino Acid Sequences of Unknown Three Dimensional Structure in a Database of Known Protein Conformations," *Proteins: Structure, Function, and Genetics* **13**, 258–271 (1992).
39. A. Godzik, A. Kolinski, and J. Skolnick, "Topology Fingerprint Approach to the Inverse Folding Problem," *Journal of Molecular Biology* **227**, 227–238 (1992).
40. S. H. Bryant and C. E. Lawrence, "An Empirical Energy Function for Threading Protein Sequences Through Folding Motifs," *Proteins: Structure, Function, and Genetics* **16**, 92–112 (1993).
41. D. Jones and J. Thornton, "Protein Fold Recognition," *Journal of Computer-Aided Molecular Design* **7**, 439–456 (1993).
42. M. J. Sippl, "Knowledge-Based Potentials for Proteins," *Current Opinion in Structural Biology* **5**, 229–235 (1995).
43. D. T. Jones, "Protein Structure Prediction in the Postgenomic Era," *Current Opinion in Structural Biology* **10**, 371–379 (2000).
44. R. H. Lathrop, "The Protein Threading Problem with Sequence Amino Acid Interaction Preferences Is NP-complete," *Protein Engineering* **7**, 1059–1068 (1994).
45. M. Wilmanns and D. Eisenberg, "Inverse Protein Folding by the Residue Pair Preference Profile Method: Estimating the Correctness of Alignments of Structurally Compatible Sequences," *Protein Engineering* **8**, No. 7, 627–639 (1995).
46. D. T. Jones, W. R. Taylor, and J. M. Thornton, "A New Approach to Protein Fold Recognition," *Nature* **358**, 86–89 (1992).
47. S. H. Bryant and S. F. Altschul, "Statistics of Sequence-Structure Threading," *Current Opinion in Structural Biology* **5**, 236–244 (1995).
48. C. Chothia, "One Thousand Folds for the Molecular Biologist," *Nature* **357**, 543–544 (1992).
49. A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia, "SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures," *Journal of Molecular Biology* **247**, 536–540 (1995).
50. C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton, "CATH—A Hierarchic Classification of Protein Domain Structures," *Structure* **5**, No. 8, 1093–1108 (1997).
51. G. T. Montelione and S. Anderson, "Structural Genomics: Keystone for a Human Proteome Project," *Nature Structural Biology* **6**, 11–12 (1999).
52. S. H. Kim, "Shining a Light on Structural Genomics," *Nature Structural Biology* **5**, 643–645 (1998).
53. T. Gaasterland, "Structural Genomics Taking Shape," *Trends in Genetics* **14**, 135 (1998).
54. S. E. Brenner and M. Levitt, "Expectations from Structural Genomics," *Protein Science* **9**, 197–200 (2000).
55. D. Fischer, "Rational Structural Genomics: Affirmative Action for ORFans and the Growth in Our Structural Knowledge," *Protein Engineering* **12**, No. 12, 1029–1030 (1999).
56. C. A. Orengo, A. E. Todd, and J. M. Thornton, "From Protein Structure to Function," *Current Opinion in Structural Biology* **9**, 374–382 (1999).
57. In practice, the term "*ab initio* method" includes a collection of different methods, dealing with different aspects of a protein's structure such as secondary structure prediction (e.g., Reference 58), prediction of contacts between amino acids (e.g., Reference 59), overall packing of the protein's secondary elements, and hybrid methods combining differ-

ent aspects of the above.[43,60–62] That is, structure prediction methods that predict any aspect of protein structure and do not make use of complete, known 3-D structures are considered to be *ab initio* methods.

58. B. Rost, "PHD: Predicting One-Dimensional Protein Structure by Profile-Based Neural Networks," *Methods in Enzymology* **266**, 525–539 (1996).
59. U. Gobel, C. Sander, R. Schneider, and A. Valencia, "Correlated Mutations and Residue Contacts in Proteins," *Proteins: Structure, Function, and Genetics* **18**, No. 4, 309–317 (1994).
60. A. R. Ortiz, A. Kolinski, and J. Skolnick, "Fold Assembly of Small Proteins Using Monte Carlo Simulations Driven by Restraints Derived from Multiple Sequence Alignments," *Journal of Molecular Biology* **277**, 419–448 (1998).
61. K. T. Simons, C. Kooperberg, E. Huang, and D. Baker, "Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences Using Simulated Annealing and Bayesian Scoring Functions," *Journal of Molecular Biology* **268**, 209–225 (1997).
62. K. T. Simons, I. Ruczinki, C. Kooperberg, B. A. Fox, C. Bystroff, and D. Baker, "Improved Recognition of Native-like Structures Using a Combination of Sequence-Dependent and Sequence-Independent Features of Proteins," *Proteins: Structure, Function, and Genetics* **34**, 82–95 (1999).
63. D. J. Osguthorpe, "*Ab Initio* Protein Folding," *Current Opinion in Structural Biology* **10**, 146–152 (2000).
64. M. Levitt and A. Warshel, "A Computer Simulation of Protein Folding," *Nature* **253**, 694–698 (1975).
65. D. Hinds and M. Levitt, "A Lattice Model for Protein Structure Prediction at Low Resolution," *Proceedings of the National Academy of Sciences (USA)* **89**, 2536–2540 (1992).
66. J. T. Pedersen and J. Moult, "Protein Folding Simulations with Genetic Algorithms and a Detailed Molecular Description," *Journal of Molecular Biology* **269**, 240–259 (1997).
67. Z. Sun, X. Xia, O. Guo, and D. Xu, "Protein Structure Prediction in a 210-type Lattice Model: Parameter Optimization in the Genetic Algorithm Using Orthogonal Array," *Journal of Protein Chemistry* **181**, 39–46 (1999).
68. K. T. Simons, R. Bonneau, I. Ruczinski, and D. Baker, "*Ab Initio* Protein Structure Predictions of CASP III Targets Using ROSETTA," *Proteins: Structure, Function, and Genetics* Supplement **3**, 171–176 (1999).
69. J. Lee, A. Liwo, D. R. Ripoll, J. Pillardy, and H. A. Scheraga, "Calculation of Protein Conformation by Global Optimization of a Potential Energy Function," *Proteins: Structure, Function, and Genetics* Supplement **3**, 204–208 (1999).
70. J. Moult, "Comparison of Potential and Mechanical Forcefields," *Current Opinion in Structural Biology* **7**, 194–199 (1997).
71. J. T. Pedersen and J. Moult, "*Ab Initio* Structure Prediction for Small Polypeptides and Protein Fragments Using Genetic Algorithms," *Proteins: Structure, Function, and Genetics* **23**, 454–460 (1995).
72. R. Unger and J. Moult, "Genetic Algorithms for Protein Folding Simulations," *Journal of Molecular Biology* **231**, 75–81 (1993).
73. R. Srinivasan and G. D. Rose, "LINUS: A Hierarchic Procedure to Predict the Fold of a Protein," *Proteins: Structure, Function, and Genetics* **22**, 81–99 (1995).
74. A. R. Ortiz, A. Kolinski, and J. Skolnick, "Nativelike Topology Assembly of Small Proteins Using Predicted Restraints in Monte Carlo Folding Simulations," *Proceedings of the National Academy of Sciences (USA)* **95**, 1020–1025 (1998).
75. B. Park, E. Huang, and M. Levitt, "Factors Affecting the Ability of Energy Functions to Discriminate Correct from Incorrect Folds," *Journal of Molecular Biology* **266**, 831–846 (1997).
76. J. Lee, A. Liwo, and H. A. Scheraga, "Energy-Based *De Novo* Protein Folding by Conformational Space Annealing and an Off-Lattice United-Residue Force Field: Application to the 10–55 Fragments of Staphylococcal Protein A and to apo calbindin D9K," *Proceedings of the National Academy of Sciences (USA)* **96**, 2025–2030 (1999).
77. J. Moult, J. T. Pedersen, R. Judson, and K. Fidelis, "A Large-Scale Experiment to Assess Protein Structure Prediction Methods," *Proteins: Structure, Function, and Genetics* **23**, ii–iv (1995).
78. J. Moult, T. Hubbard, S. H. Bryant, K. Fidelis, and J. T. Pedersen, "Critical Assessment of Methods of Proteins Structure Prediction (CASP): Round II," *Proteins: Structure, Function, and Genetics* Supplement **1**, 2–6 (1997).
79. D. T. Jones, "Progress in Protein Structure Prediction," *Current Opinion in Structural Biology* **7**, 377–387 (1997).
80. D. Fischer, "Modeling Three-Dimensional Protein Structures for Amino Acid Sequences of the CASP3 Experiment Using Sequence-Derived Predictions," *Proteins: Structure, Function, and Genetics* Supplement **3**, 61–65 (1999).
81. D. Shortle, "Structure Prediction: Folding Proteins by Pattern Recognition," *Current Biology* **7**, R151–R154 (1997).
82. R. L. Dunbrack, D. L. Gerloff, M. Bower, X. Chen, O. Lichtarge, and F. E. Cohen, "Meeting Review: The Second Meeting on the Critical Assessment of Techniques for Protein Structure Prediction (CASP2)," Asilomar, CA (December 13–16, 1996); *Folding & Design* **2**, No. 2, R27–R42 (1997).
83. D. Fischer, "Hybrid Fold Recognition: Combining Sequence Derived Properties with Evolutionary Information," *Proceedings of the 1st Pacific Symposium on Biocomputing* (2000), pp. 119–130.
84. L. A. Kelley, R. M. MacCallum, and M. J. E. Sternberg, "Recognition of Remote Protein Homologies Using Three-Dimensional Information to Generate a Position Specific Protein Matrix in the Program 3D-PSSM," *RECOMB99—Proceedings of the Third Annual Conference on Computational Biology*, S. Istrail, P. Pevzner, and M. Waterman, Editors, Association for Computing Machinery, New York (1999), pp. 218–225.
85. S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs," *Nucleic Acids Research* **25**, No. 17, 3389–3402 (1997).
86. A. Krogh, M. Brown, I. S. Mian, K. Sjolander, and D. Haussler, "Hidden Markov Models in Computational Biology: Applications to Protein Modeling," *Journal of Molecular Biology* **235**, 1501–1531 (1994).
87. A. R. Ortiz, A. Kolinski, P. Rotkiewicz, B. Ilkowski, and J. Skolnick, "*Ab Initio* Folding of Proteins Using Restraints Derived from Evolutionary Information," *Proteins: Structure, Function, and Genetics*, Supplement **3**, 177–185 (1999).
88. D. J. Osguthorpe, "Improved *Ab Initio* Predictions with a Simplified, Flexible Geometry Model," *Proteins: Structure, Function, and Genetics* Supplement **3**, 186–193 (1999).
89. Y. Samudrala, R. Xia, E. Huang, and M. Levitt, "*Ab Initio* Protein Structure Prediction Using a Combined Hierarchical Approach," *Proteins: Structure, Function, and Genetics* Supplement **3**, 194–198 (1999).

90. A. L. Lomize, I. D. Pogozheva, and H. I. Mosberg, "Prediction of Protein Structure: The Problem of Fold Multiplicity," *Proteins: Structure, Function, and Genetics* Supplement **3**, 199–203 (1999).
91. M. J. E. Sternberg, P. A. Bates, L. A. Kelley, and R. M. MacCallum, "Progress in Protein Structure Prediction: Assessment of CASP3," *Current Opinion in Structural Biology* **9**, 368–373 (1999).
92. P. Koehl and M. Levitt, "A Brighter Future for Protein Structure Prediction," *Nature Structural Biology* **62**, 108–111 (1999).
93. C. Venclovas, A. Zemla, K. Fidelis, and J. Moult, "Some Measures of Comparative Performance in the Three CASPs," *Proteins: Structure, Function, and Genetics* Supplement **3**, 231–237 (1999).
94. M. J. Sippl, P. Lackner, F. S. Domingues, and W. A. Koppersteiner, "An Attempt to Analyze Progress in Fold Recognition from CASP1 to CASP3," *Proteins: Structure, Function, and Genetics* Supplement **3**, 226–230 (1999).
95. A. Marchler-Bauer and S. H. Bryant, "A Measure of Progress in Fold Recognition?" *Proteins: Structure, Function, and Genetics* Supplement **3**, 218–225 (1999).
96. H. M. Berman, J. Westbrook, Z. Feng, G. Gillil, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Research* **28**, 235–242 (2000).
97. J. M. Bujnicki, A. Elofsson, D. Fischer, and L. Rychlewski, "LiveBench: Continuous Benchmarking of Protein Structure Prediction Servers," *Protein Science* **10,** 352–361 (2001).
98. N. Siew, A. Elofsson, L. Rychlewski, and D. Fischer, "Max-Sub: An Automated Measure for the Assessment of Protein Structure Prediction Quality," *Bioinformatics* **16**, 776–785 (2000).
99. D. Fischer, A. Elofsson, and L. Rychlewski, "The 2000 Olympic Games of Protein Structure Prediction," *Protein Engineering* **13**, 667–670 (2000).
100. D. Butler, "IBM Promises Scientists 500-fold Leap in Supercomputing Power," *Nature* **402**, 705–706 (1999).
101. F. Allen et al., "Blue Gene: A Vision for Protein Science Using a Petaflop Supercomputer," *IBM Systems Journal* **40**, No. 2, 310–327 (2001, this issue).
102. On December 3, 2000, the CASP4/CAFASP2 meeting was held in Asilomar, California. Predictions based on the automated results from various fold-recognition servers and filed under the group name "CAFASP-CONSENSUS" scored within the top 7 performing human groups in CASP4, as judged by the CASP4 assessor. For more details, see the CASP4 and CAFASP2 home pages (listed in the summary and discussion of this paper) and the upcoming special issue of the journal *Proteins: Structure, Function, and Genetics*.
103. Parts of the CAFASP and LiveBench material presented in this paper were adapted from Reference 99.

Accepted for publication January 5, 2001.

Naomi Siew *Department of Chemistry, Ben Gurion University, Beer-Sheva 81405, Israel (electronic mail: nomsiew@cs.bgu.ac.il).* Ms. Siew received a bachelor's degree in chemistry in 1994 and a *cum laude* master's degree in pharmaceutical chemistry from the Hebrew University of Jerusalem in 1996. In 1998 she worked at the Scripps Research Institute in La Jolla, California, under the supervision of Jacquelyn Fetrow and Jeffrey Skolnick. Since 1999 she has been a graduate student at Ben Gurion University under the joint supervision of Joel Bernstein from the Chemistry Department and Daniel Fischer from the Bioinformatics/Computer Science Department. Her current work focuses on sequence ORFans in whole genomes.

Daniel Fischer *Bioinformatics/Computer Science Department, Ben Gurion University, Beer-Sheva 81405, Israel (electronic mail: dfischer@cs.bgu.ac.il).* Dr. Fischer is an assistant professor at Ben Gurion University where he leads the Bioinformatics Group. His work focuses on protein structure prediction, computational structural biology, computational analysis and interpretation of genomes, and bioinformatics in general. Dr. Fischer received his doctoral degree in computer science from Tel Aviv University in 1994. He has been a successful participant of the CASP, CAFASP, and LiveBench experiments. Since 1975 he has been a computer-chess aficionado, but the last time he beat a computer-chess program was in 1976.