Published in final edited form as:

J Struct Funct Genomics. 2012 December ; 13(4): 213–225. doi:10.1007/s10969-012-9146-2.

Assessing the Accuracy of Template-Based Structure Prediction Metaservers by Comparison with Structural Genomics Structures

Dominik Gront^{1,2,3}, Marek Grabowski^{1,3}, John Raynor^{1,3}, Karolina L. Tkaczuk^{1,3}, and Wladek Minor^{1,3}

¹Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA ²University of Warsaw, Faculty of Chemistry, Pasteura 1, 02-093 Warsaw, Poland ³Midwest Center for Structural Genomics

Abstract

The explosion of the size of the universe of known protein sequences has stimulated two complementary approaches to structural mapping of these sequences: theoretical structure prediction and experimental determination by structural genomics (SG). In this work, we assess the accuracy of structure prediction by two automated template-based structure prediction metaservers (genesilico.pl and bioinfo.pl) by measuring the structural similarity of the predicted models to corresponding experimental models determined *a posteriori*. Of 199 targets chosen from SG programs, the metaservers predicted the structures of about a fourth of them “correctly.” (In this case, “correct” was defined as placing more than 70% of the alpha carbon atoms in the model within 2 Å of the experimentally determined positions.) Almost all of the targets that could be modeled to this accuracy were those with an available template in the Protein Data Bank (PDB) with more than 25% sequence identity. The majority of those SG targets with lower sequence identity to structures in the PDB were not predicted by the metaservers with this accuracy. We also compared metaserver results to CASP8 results, finding that the models obtained by participants in the CASP competition were significantly better than those produced by the metaservers.

Introduction

Following the advances in gene sequencing technology in the 1990s, the size of the known protein sequence universe – the set of all known polypeptide sequences – has expanded exponentially [1, 2], posing a challenge for structural biology to structurally and functionally characterize this space. The relatively high cost and difficulty of experimental determination of protein structure stimulated development of two very different approaches to address this challenge: computational structure prediction and structural genomics (SG).

Corresponding authors: Marek Grabowski, marek@iwonka.med.virginia.edu, Wladek Minor, wladek@iwonka.med.virginia.edu, Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA.

In theory, the 3-D structure of a protein should be predictable solely from its amino acid sequence [3]. Such prediction is conceptually appealing but extremely difficult in practice. Current *de novo* structural prediction methods are unreliable, producing reasonable results only for relatively short polypeptide sequences (up to 85 amino acids) [4].

Another group of prediction methods, template-based or comparative modeling, uses the experimentally determined structure of a sequence relative (e. g. a homolog) as a template for reconstruction of the 3D structure of the investigated (“target”) protein. In this work, we investigate the accuracy of two automated tools for template-based structure prediction.

Initially, the ultimate goal of worldwide SG projects was structural characterization of a large subset of the protein universe, so that a structure of nearly any protein would be available either directly from the Protein Data Bank (PDB) or through computational methods (homology modeling). A major step toward this goal was the creation of the NIH-sponsored Protein Structure Initiative (PSI). It was hoped at the outset of PSI that by solving structures of 16,000 carefully selected proteins, it would be possible to obtain structures for nearly 90% of all proteins [5]. While particular SG centers adopted various target selection strategies, the common requirement was that new targets should have sequence similarity below 30% to existing experimentally determined structures. This cutoff was a conservative estimate: a vast majority of pairs of sequences sharing 30% or greater sequence identity are homologous, with few false positives. Thus, for a protein with more than 30% sequence identity to a known structure, a good quality theoretical model can usually be built using the existing structure as a modeling template. This potentially could eliminate the need for experimental structure determination.

In this work we analyze and compare the performance of automatic methods for structure prediction based on a set of protein sequences whose structures have also been obtained experimentally. As a dataset we used proteins that were experimentally solved and deposited by SG centers, most of which were produced by the four large-scale Protein Structure Initiative (PSI) centers in the period of April to August 2008. As a general rule, new structures deposited to the PDB by PSI centers were released to the public within two to three weeks. This delay provided us a window of time to conduct an unbiased experiment; i.e., building a predicted model without running the risk of the automatic servers choosing the experimental structure of the protein as a template for itself.

The goal of our computational experiment was to answer the following questions:

- i. Could the structures of some of these targets have been predicted using fully automatic tools, without experimental structure determination?
- ii. What is the quality of the predicted models generated?
- iii. How does the model quality depend on sequence similarity to the template structures?
- iv. Is the 30% sequence identity cutoff a sensible one?

Our computational experiment is conceptually similar to the ongoing Critical Assessment of Protein Structure Prediction (CASP) competition [6]. Our goal was to evaluate the reliability

of 3D structures which can be readily obtained by non-experts (i.e. biologists without extensive structural modeling experience) through publicly available, easy-to-use prediction metaservers. In other words, we are comparing the automated, “high-throughput” structure prediction pipeline provided by these metaservers with the high-throughput experimental SG pipeline.

In template-based structure prediction, the target sequence is used to search for potential template structures. Templates are detected by algorithms based on sequence alignment, profile alignment or threading. As a result of this step one gets an alignment between a target sequence and a template structure. The alignment provides a one-to-one correspondence between residues in a target and template proteins. By aligning two proteins by primary sequence, it is assumed that sequence similarity in a region implies structural similarity there as well. Modeled structure fragments are extracted and the missing parts are built by *de novo* modeling. The final model may be subjected to overall structure refinement and scoring.

Due to the enormous number of possible conformations for a given polypeptide chain, the problem of accurately predicting a 3-D protein structure, even using a template, is still far from being solved. The MODELLER program [7], used in this work, is slightly less accurate than other tools for template-based modeling [8], but is fully automated and integrated with the metaservers analyzed in this work. MODELLER takes a target-to-template sequence alignment as well as a target protein structure as an input. From this data the program derives spatial restraints, and a scoring function based on the restraints and a statistical force field (trained on structures already deposited into PDB) is optimized.

The Bioinformatics Links Directory (http://bioinformatics.ca/links_directory), which features curated links to molecular resources, tools and databases, lists more than 70 servers related to protein structure prediction and modeling. Which server, out of the dozens that identify template models using different algorithms, is likely to yield the best results? No single server or algorithm is likely to predict the best model for every target; metaservers can be used to generate the best possible results [9]. A metaserver is a frontend to many prediction servers spread over the Internet. When a user submits a task to a metaserver, the task is forwarded to several prediction servers. When the servers return their results, the metaserver formats them in a unified way, calculates a consensus score and displays it on a web page. Ideally, to prevent bias, the servers polled by the metaserver should calculate results independently from one another, but in practice this condition may not always be satisfied.

In our computational experiment, we used metaservers to predict structures of proteins targeted by the SG centers. Sequences of SG targets are available before their structures are experimentally determined in the TargetDB database [10]. This enabled us to submit sequences of targets to metaservers and perform structural predictions before the structure of the target itself was released by the PDB, which prevented their unintended use as a potential template by the prediction servers. In the experiment we performed fully automatic predictions, meaning that we did not improve the alignments between the target and the template manually, either to select better templates or to refine obtained models. We also did

not correct for shortcomings in software, nor used particular prediction servers directly. As previously mentioned, we sent sequences to the metaservers and collected the results produced, emulating the way in which researchers without an extensive background in structural modeling would use the metaservers.

Materials and Methods

Target sequences

We attempted to predict protein structure for 216 targets from the SG programs deposited into the PDB in the 16-week span from April 21, 2008 to July, 30, 2008. The targets were divided into two groups: those solved by NMR or by X-ray crystallography with phases determined experimentally (200 structures) and targets solved by molecular replacement (MR; 16 structures). The structures solved by MR were excluded from analysis, because by definition each already has a known template. Of the remaining structures, one of the deposits (2JS6) was withdrawn and was also excluded, reducing our dataset to 199 targets.

The sequences of the targets were based on XML files which the SG centers posted on their web sites and submitted to the TargetDB database [10] and were harvested into a relational database. Sequences of the proteins were submitted to the prediction metaservers just after the targets were solved and deposited to the PDB, but before the coordinates of the deposits had been released. Most of the targets were determined by the four large-scale PSI centers: the Northeastern Center for Structural Genomics (NESGC) – 62, the Joint Center for Structural Genomics (JCSG) – 55, the Midwest Center for Structural Genomics (MCSG) – 51, and the New York Structural Genomix Research Center (NYSGXRC) – 45. Some were solved by the Montreal-Kingston Bacterial Structural Genomics Initiative (BSGI) – 7 and by the Structure to Function project – 1.

Modeling methods

We calculated predictions using two widely used metaservers: bioinfo.pl [11] and genesilico.pl [12]. Each of these tools sends a query sequence to a number of prediction servers and creates a consensus of the obtained results. The two metaservers differ in the set of servers employed and in the method by which a consensus is formed: bioinfo.pl uses the 3d jury algorithm [11], while the genesilico.pl uses Pcons5 [13]. As the final result, each metaserver returns a list of the 10–20 best models as ranked by the consensus scoring. In total we gathered and analyzed 4944 models. We recorded target-to-template alignments and 3D coordinates of the models reported by the metaservers.

Domain parsing

To evaluate modeling results on separate domains, each target structure was split into domains by means of the Protein Domain Parsing (PDP) algorithm [14].

Evaluation of the metaserver output

Each target protein used as an input to our modeling procedure was characterized by a single amino acid sequence composing one polypeptide chain, thus the corresponding models returned by the prediction servers were monomeric. However, in many cases experimentally

determined structures are oligomers containing more than one copy of a modeled chain in the asymmetric unit (in the case of structures determined by X-ray crystallography). In the evaluation of the metaserver output, we selected one chain as a reference according to the following rules:

- The chain covering the longest part of a target sequence was chosen.
- If two or more chains cover the same length of a target sequence, the chain with the lowest average B-factor was chosen.

The selection of reference chain had a considerable influence on the modeling results. An extreme example is structure 3CZB, which according to the SEQRES records in the PDB file nominally comprises 2 identical chains of 351 residues each. The sequences of both chains, however, contain a few point mutations and both chains have disordered (unmodeled) regions, resulting in a relatively large crmsd (coordinate root mean square deviation) of 6.3Å between the two chains.

GDT and crmsd measures

In most cases the target sequence for prediction differed from (i.e. was usually longer than the sequence extracted from the corresponding PDB deposit. This was often a consequence of disordered regions of the structure and resulted in an inability to reliably model poor or missing regions of the electron density map (in the case of X-ray diffraction structures). In a few cases, inclusion of cloning artifacts like affinity tags in the structure resulted in the sequence in the structure being longer than the original sequence. Therefore, prior to any quality assessment, the original target sequence was aligned with the observed sequence in the deposited structures. All quality parameters (described below) were calculated on the subset of residues consisting of the intersection of the two sequences. N is defined as the number of atoms in this subset.

For each model we calculated its crmsd distance to the native structure, defined as:

$$1/N \sqrt{\sum (x_j^m - x_j^n)^2 + (y_j^m - y_j^n)^2 + (z_j^m - z_j^n)^2},$$

where x_j^m , y_j^m and z_j^m are the coordinates of the j -th atom in the model, and x_j^n , y_j^n and z_j^n are the coordinates of the corresponding atom in the native structure, and the sum is taken over all N atoms after the optimal superposition is calculated [15]. This parameter, although frequently used, is not a good measure for model quality. First, crmsd depends on the number of atoms taken into the calculation (i.e. the length of the polypeptide chain). This weaker than linear dependence cannot be expressed analytically, and therefore there is no direct way to rescale crmsd values to a length-independent measure. The lengths of our modeling targets span a wide range, from 55 to 983 amino acids. Moreover, quantitative comparison of crmsd values makes sense only in the range of small values: clearly a 0.5Å crmsd model is more accurate than a 2.5Å one, but relative accuracy is more difficult to judge for models with 10Å and 50Å crmsd, although in both cases one model has a crmsd 5 times smaller than the other.

To address this shortcoming, an alternative method for determining predicted model quality was used: the GDT (Global Distance Test). GDT measures what fraction of one structure can be superimposed on another within a given inter-atomic distance (as a dimensionless fraction from 0.0 to 1.0). For example $GDT(2\text{\AA}) = 0.7$ means that when 70% of a model structure has been optimally aligned and superimposed on a reference structure, any atom from that fragment is closer to its counterpart than $2\text{\AA}$. The selected atoms do not need to form a continuous segment. In the case of partial models, crmsd values are lower and GDT values higher than they would be for full-length models. To be able to accurately compare GDT values for all models returned by the two metaservers, we normalized the GDT fraction by the number of ordered residues in the PDB deposit rather than by the number of residues in a calculated model, which means that unmodeled loops and fragments in the *a posteriori* structure would be excluded from the calculations. All crmsd and GDT values were calculated with the `jbcl.calc.structural` module of the BioShell package [16, 17].

Sequence identity measures

We used two different parameters to assess target sequence identity with respect to available templates. The first one, called the *alignment sequence identity* (ASI) for a predicted model, is defined as the number of identical residues in an alignment of the target and the template used to build the model, divided by the length of the target (modeled) sequence. The second, the *reference sequence identity* (RSI) for a predicted model, is defined as the sequence identity between the target and the template structure with the most accurate template (according to highest GDT or crmsd to the subsequently determined experimental structure) divided by the length of the target protein. Note this means that the RSI for a predicted model is not necessarily based on the same template as the one used to generate the prediction. For the PSI targets, an important metric used to determine the novelty of the deposits is the highest relative sequence identity to any of the proteins available in the PDB at the moment of deposition of the target. This concept is roughly equivalent to the RSI—the two measures would be completely identical if all prior deposits were known to the metaservers at the time of deposition of the target.

Alignments were calculated by BioShell package using local dynamic programming [18] with an affine gap penalty (using the BLOSUM50 substitution matrix [19], a gap opening penalty of -10 , and a gap extension penalty of -2). These values are the same as those used for target selection (or exclusion) by the MCSG SG center.

Results

To evaluate the results from our computational experiment, we measured the structural similarity between models generated by the modeling metaservers and the corresponding subsequent structures solved experimentally. The similarity was measured using both the GDT and crmsd measures (defined in *Materials and Methods* above). Taking into account that predicted model quality strongly depends on sequence similarity between a target and a template, and the fact the metaservers may output models comprising subsequences of the same target differing in length, we used both the alignment sequence identity (ASI) and

reference sequence identity (RSI) to measure sequence similarity (described in more detail above).

Fig. 1 shows the sequence similarity of the predicted models with the experiment structures as a function of the two sequence identity measures. Fig. 1A and B show that predicted model quality increases—as measured by decreased crmsd (Fig 1A) or increased GDT(1Å) (Fig 1B)—as ASI increases. Similarly, Fig. 1C and D show that predicted model quality also increases as RSI increases. Each of these plots display data points corresponding to 4944 models of 199 targets. The points in blue identify the best models (as measured by similarity to the experimentally determined structures) created by the metaservers. Obviously, without knowledge of the experimental structure, it is not possible to identify the best model. The metaservers attempt to score the predicted models according to their own criteria (marked in green on Fig. 1C and D). In most cases the highest scoring model was worse than the best model identified *a posteriori*. Only for 36 targets (18.1%) was the best model in terms of GDT(2Å) ranked as the highest-scoring one by either genesilico.pl [12] or bioinfo.pl [11].

Each vertical strip visible in Fig. 1C and D corresponds to a single target sequence and depicts the variation between different modeling results (likely based on different templates) for this target. For example, there are four vertical stripes overlaid on Fig. 1D with RSI close to 1.0 (which means that the target and template sequences were nearly identical). For three of them, the best models correspond to very good modeling results: GDT(2Å) is above 0.9 in two cases and 0.7 in the third. The other predicted models for these target sequences cover almost the whole range of GDT(2Å), which were built on templates worse than the optimal one. In the fourth case the most accurate predicted model is of relatively poor quality (GDT(2Å) = 0.48), even though the sequence of the target and the template structure are identical. This is discussed in more detail below.

A comparison of the results from the two metaservers must take into account the varying degree of template coverage of targets, as each server uses a different approach to handle gaps in a target-template sequence alignment. Bioinfo.pl uses the MODELLER [7] program to create all-atom structures that cover the whole sequence of a target. MODELLER tries to reconstruct the coordinates of the residues that do not have their counterparts in a template and were not established experimentally. Therefore models returned by bioinfo.pl cover 100% of the target sequence. In contrast, genesilico.pl by default applies its own modeling method, essentially copying the coordinates of atoms directly from the template and omitting the parts of the target that have no counterparts in the template. As a result, models produced by genesilico.pl generally cover only parts of the target sequence (in a few cases, only 10% of the target).

While MODELLER is generally able to correctly reconstruct short loops (of 20 amino acids or fewer), when there is a longer missing fragment, it sometimes inserts an unstructured, randomly-oriented polypeptide chain, resulting in a large value of crmsd between the model and the experimental structure. In such cases, crmsd is a poor measure of the modeling accuracy, as it overemphasizes the errors of the interpolated sections. For example, the model of 3CKW built using 3COM as a template by the bioinfo.pl server had a very poor crmsd value (>33 Å) to the *a posteriori* 3CKW structure, while the value of GDT(2Å) was

0.88. This indicates that 88% of the structure was modeled with accuracy better than 2 Å, a reasonably good fit.

26% of the SG structures could be predicted with 70% of atoms in correct positions

For 55 of the 199 targets, the best model found by the metaservers had $GDT(2\text{\AA}) > 0.7$, meaning that 70% of the atoms in the model could be placed within 2 Å of the positions determined experimentally. For almost all of these targets (53), the RSI was greater than 25%, meaning that the metaservers had found a template in the PDB with greater than 25% sequence identity to the target.

When sequence identity between a target and its template decreases, the modeling process must introduce more changes into the template to build a model: the missing parts must be rebuilt and proper side chains added. When sequence similarity drops below a threshold—originally suggested to be about 25% identity for alignments >80 amino acids long [20]—it is very difficult to determine if two sequences are structurally homologous using sequence information alone, and template-based modeling is in general no longer possible. However, out of 147 targets which had no template in the PDB above 30% sequence identity (the cutoff used by most SG programs), 17 targets yielded predicted models with $GDT(2\text{\AA}) > 0.7$.

Templates with >30% sequence identity permit reasonably good modeling

Our results show that in most cases, when a template can be found with sequence identity more than 30%, $GDT(2\text{\AA})$ for the best model exceeds 0.8 (Fig 1D). 51 out of the 199 SG targets (25.6%) are >30% identical to their best template (these are given in Table 1). Moreover, 16 of the targets are >40% identical to their best template. In some cases, these targets represent proteins of biomedical interests, or those requested by community, for which structure determination was desired regardless of sequence similarity to proteins of known structure. Some of the other cases likely represent “structural coverage” targets that were previously solved by other structural biologists.

The effect of template choice on modeling accuracy

As mentioned above, the metaservers are only frontends to other servers that perform the actual calculations. Ideally each server would search through the whole PDB database to find the best possible template for modeling; however, we found that some servers appeared to only search a subset of the whole PDB. The servers polled by the metaservers differ both in scoring system and the set of structures (as described above) used for template selection and frequently selected different structures as the best template. In general, when the sequence similarity between the sequences of the target and template was relatively high, the resulting models were relatively accurate. At lower levels of sequence identity, choosing a template capable of yielding an accurate modeling was much more difficult. Several templates (typically 6) were selected for many targets in our experiment (Fig 2). In a majority of cases, selecting a poor template resulted in much more inaccurate models than errors in sequence alignment or model building (Fig. 2), indicating the importance of proper template selection in generating an accurate template-based model.

Identical sequences do not always predict identical structures

When a template sequence was identical to the target sequence, one might expect the model predicted by the metaservers, the template and the experimental structure to be all identical. However, this was not always the case. For some of these models, the GDT(2Å) between the model and the native structure was significantly less than 1.0. The reasons for these discrepancies include:

- *Differences due to experimental structure determination, either in the template or in the target structure.* In some cases the experimental structure of a modeled target has disordered regions that are differently oriented in the template (e. g. 3D3N vs. 3BXP)
- *Different relative domain orientations.* PDB deposit 2OZN comprises two SG targets: Cohesin (chain A) and Fivardoc (chain B). The sequence of chain A of 2OZN is identical to that of 2JH2; and the sequence of chain B of 2OZN is identical to that of 2JNK (which was solved by NMR). In the 2OZN heterodimer, the structures of the two targets differ from the corresponding monomer structures. In particular the NMR structure of 2JNK, which comprises two helical domains, differs greatly from its counterpart, the chain B of 2OZN_B, in which the two domains are arranged differently with respect to one another, resulting in crmsd = 2.7 Å and GDT(2Å) = 0.46. Another example is the pair of 3D4O and 2RIR, where the relative spatial arrangements of the two domains of the identical polypeptides are slightly different. This relatively small change affects structure similarity measures based on structural superimposition of the two proteins, resulting in a relatively low GDT(1Å) = 0.48.
- *Differences in loop regions.* E.g. the pair of 3DBI and 3BRQ.
- *Different experimental methods for template and target.* Target 2OZN (subsequently solved by X-ray crystallography) was modeled on the 2JNK protein solved by NMR. Although the target and template have identical sequences, due to different domain orientation, discussed above, the GDT(1Å) for the best model is only 0.22.

In general, polypeptide chains that were identical in sequence (as listed the SEQRES records of the PDB files) in many cases differed significantly in structure. For 386 pairs of identical chains extracted from 75 experimental structures, the mean GDT(1Å) between the identical chains was 0.81 (median 0.91). For GDT(2Å) the mean and median were 0.91 and 0.98, respectively. In a few cases, the differences are substantial. For example, the crmsd distance between chains A and B of 3CQB is 5.2 Å, with GDT(1Å) = 0.70 and GDT(2Å) = 0.83. In this case, the discrepancy is due to unordered N-terminal regions (Fig. 3.).

Discussion

The impact of the PSI on template-based modeling

As a good template structure is the most important factor in the success of template-based modeling, one of the main goals of the PSI was to provide templates for as many target sequences without homologs as possible. We conducted our computational experiment in

the 8th year of the PSI, when SG centers had already deposited over 2500 structures into the PDB, and thus the modeling should have benefitted from the use of PSI structures as templates. Indeed, out of 4944 models gathered during the experiment, 1963 (40.9%) used a structure deposited by a PSI center as a template. Moreover, 1054 models (22.0%) were based on a structure deposited by a different center than the one that eventually deposited an experimental structure of the target. When only the model with the highest GDT(1Å) for each target was considered, 96 out of 199 models (48.2%) used another PSI target as a template, and 59 (29.6%) used a PSI template from a different center. This conclusion is even more interesting when one takes into account that all PSI centers are tasked to solve novel proteins with no more than 30% sequence identity to proteins of known structure.

The modeling tools yield models that do not significantly differ from the template

A majority of the servers do not in fact alter the template much, and essentially return the original template structure as the final predicted model. In our work we used the MODELLER program, which rebuilds the parts of the target sequence that are missing in the template and optimizes the model in a full-atom force field. Other atomic positions are altered only to remove steric clashes. Therefore structural changes are usually limited to loop regions. An *a posteriori* analysis based on a comparison between calculated models and the relevant template regions confirms that the models are very close to the templates, always with <2 Å crmsd for the aligned fragments (data not shown). Along with the results shown in Fig. 2, this suggests that proper template selection is the most critical factor in generating an accurate template-based model.

No single server produced significantly more accurate results than the others

We collected data from 13 servers, as collated by the two metaservers. When the results are compared, no single server produced significantly more accurate results than the others (Table 2), though the servers differed in the number of accurately predicted targets. This may be due to real differences in the effectiveness of the template selection and modeling algorithms used, but may also be due to other factors beyond the control of each individual server. These factors include:

- A server may produce an accurate model of a target, but the metaserver did not rank the solution among the highest scoring ones.
- It might have taken too long for a server to send back a result. In this work we gathered results for only two weeks after submission, because PSI deposits are kept on hold by the PDB for only that time. Not all of the servers answered before the deadline.

A priori, one may expect that poor template assignments occur mostly when the target has low sequence similarity to existing structures in the PDB. The more similar the sequence of a potential template, the more accurate template selection should be. However, this was not always observed. For the 16 targets where a very similar template (RSI > 40%) was already present in the PDB database at the time of deposition, not a single server managed to find solutions for all 16 of these targets. The highly similar templates for each of these targets could have been found by a trivial BLAST [21] search, provided that the reference databases

used by the servers were kept up-to-date. This underscores the importance of proper maintenance of bioinformatics tools and databases, and indicates that the advantages gained by the rapid growth of the number of available templates in the PDB may be not be properly leveraged due to the lack of timely updates.

Suitability of X-ray vs. NMR-derived templates

Out of 199 of the target proteins, structures of 33 were subsequently solved by NMR (16.58%). Similarly, many of the template structures were solved by NMR rather than X-ray diffraction, which sometimes led to odd results. The example of target 2OZN (experimentally solved by X-ray diffraction), modeled using 2JNK as a template (solved by NMR), was already mentioned above. Another curious example is that of 2K3I. Both the experimental structure of 2K3I and its best modeling template 2JZ5 (sharing 36.7% identical residues) were solved by NMR. Despite the relatively high RSI, the best predicted model of 2K3I had a GDT(2Å) of only 0.58 with respect to the experimental structure. This value rose to 0.86 by excluding disordered regions.

There were 56 targets for which the metaservers returned at least one model based on an X-ray template and at least one model based on an NMR template. In the majority of these cases, the most accurate predicted model was based on an X-ray template (46 out of 56). For the 10 models where the NMR structures served as better templates, in all cases the sequence identity of the target to the sequence of the NMR template structures were higher. When two templates (one of each type) were similar in sequence identity to the target, the predicted models built using an X-ray template was always more accurate than those built with an NMR template. This applied even if the experimentally derived structure of the target was subsequently solved by NMR.

The multidomain nature of proteins is one of the biggest problems for structure modeling

Almost half of the target proteins comprise more than one domain, creating difficult problems for automated modeling tools. For many of the templates identified for multiple domain targets, only one of the domains of the template was homologous by sequence identity to the target, and thus typically only that domain was correctly modeled. Automated modeling servers usually build the full target structure based on a single template, although a more accurate model would be predicted if domains separately built using multiple template structures were combined, or if the template was constructed as an aggregate of multiple domain templates.

In cases where the template structure contains the same domains as the target, it is possible that the relative orientations of the domains to one another differ in the target versus the template. Of the 16 targets with templates with >40% RSI, 10 targets had predicted models with the correct relative domain orientations. Overall there is no correlation between correctness of relative domain orientation and sequence identity (beyond a certain threshold of sequence identity; data not shown).

When one excludes all multi-domain targets from Figure 2 (data not shown), modeling using templates with very likely homology (RSI $\geq$ 40%) always gives accurate results. Interestingly, the accuracy of predictions, when evaluated for particular domains of multi-

domain targets, was not significantly better than the accuracy of predictions for the whole length of the protein. This may be a consequence of the fact that in many cases modeling servers failed to find a good template for at least one of the domains. It is also possible that an additional part of a query sequence that cannot be aligned with the sequence of the best template confuses the template selection algorithm.

In some cases a structure deposited into the PDB does not cover the whole sequence of the protein target used for modeling. This happened relatively infrequently: roughly 30% of target sequences extracted from PDB deposits cover less than 90% of modeled sequences. In an extreme example, however, one template covered less than 20% of the target sequence.

Humans are better than metaservers

Our study coincided with the eighth edition of CASP, the Critical Assessment of Techniques for Protein Structure Prediction [22], a worldwide experiment and contest for protein structure prediction taking place every two years since 1994. Its course is similar to the work we described here. Amino acid sequences are provided to predictors; the structures remain unreleased until the end of the experiment. Despite the conceptual similarity between our experiment and the CASP experiment, there are several important differences. CASP is conducted to test the general performance of state-of-the-art prediction methods (and to motivate innovation), while our goal is to test the performance of easy-to-use publicly available services for a non-expert user who hopes to predict structures of a protein of interest.

Though CASP targets may come from sources other than SG, 81 of the targets investigated here were also CASP8 targets, allowing a comparison of the quantitative assessment of metaserver predictions described in this work with CASP8 results. The results of CASP8 were assessed by human experts (assessors) who knew both the computed models and the experimentally determined structures. They arbitrarily divided all native structures of the target protein into domains and removed flexible tails so CASP8 results are based on different protein fragments than those defined in our analysis.

We assessed the quality of all the models gathered during our experiment for the 81 relevant targets based on the domain definitions published by the CASP8 assessors (111 domains). For every target the best CASP model is at least as accurate (as measured by GDT to the known structure) as our best model. In some cases, however, the difference is not significant. Results for target T0458 (3DEX) are almost identical (results not shown). For 7 domains, the automated template-based modeling methods managed to obtain results very similar (at most 1 residue misaligned according to GDT(2Å)) to those predicted by the best human experts. However, on the average the best CASP models were more accurate (by a GDT(2Å) difference of about 0.12) than the best models found by the bioinfo.pl metaserver. Only 12 domains yielded models by the metaservers that would rank among the top 10 CASP server results. In around 30 cases the metaservers did not choose the best template and therefore produced significantly worse models. 11 out of 111 target domains fell into the *ab initio* category, as it was not possible to detect any template at all (data not shown).

There are several explanations as to why the CASP8 predictions outperformed the automated predictions:

- As described above, the metaservers do not parse the sequences into domains.
- In a number of cases the target-to-template alignment resulting from automated methods is incorrect (see [23] for an illustrative example). Manual corrections done by a human expert may greatly improve the accuracy of the final model.
- Final model selection done by a human expert is, in a majority of cases, superior to automatic selection and scoring methods.

Recent progress in the field

The results presented in this work were obtained nearly four years ago. Since then another two rounds of CASP have taken place, CASP9 and CASP10. CASP10 ends in fall 2012, so the final results are still yet to be announced. The most recent literature reports [24, 25] summarizing the results of CASP9 in 2010 indicate that protein structure modeling methods continue to advance, although at a more modest pace than in earlier CASP experiments [26], resulting in only an incremental improvement of comparative modeling performance.

Kryshtafovych et al. [25] report that significant progress has been made in template-based modeling, resulting in improved model accuracy for targets in the midrange of difficulty, as well as for short template-free modeling targets. According to the authors, improvements in template-based modeling can most likely be attributed to improved methods for combining information from multiple templates. Multiple-template-based modeling is an emerging area of modeling in which a number of innovative developments have been reported in past years. In the case of the short template-free modeling targets, the improvement is due to a greater variety of methods available, as well as improved methods for identifying the best model from a set of generated models.

To test whether our analysis of automatic structure prediction by metaservers is still relevant today, we have repeated the process for three structures solved more recently by SG centers: T0752, T0756 and T0733, which were candidates for the server-only part of the CASP10 competition. For T0752 both the metaservers and CASP were able to model most of the structure, with $GDT(2 \text{ \AA}) = 0.878$ for the CASP server and $GDTA(2 \text{ \AA}) = 0.810$ achieved by the metaservers. For the other two proteins, only partial models were built. For all three, as in our original analysis, the best servers participating in CASP had better results than the metaservers. In light of these examples, it appears likely that the accuracy of prediction by metaservers has not increased significantly since we performed the original analysis. We would expect our results, if calculated today, to remain very similar, providing that the new PDB deposits were removed from the set of considered template structures.

Two of the most important recent developments in protein structure modeling, HHSearch and I-Tasser have not been incorporated into the methods of the metaservers we studied. HHSearch, introduced by Soding [27], utilizes a hidden Markov Model method for template detection and alignment. The I-Tasser server for protein structure prediction [28] has in some cases managed to propose more accurate predictions than human experts.

In the field of *de novo* modeling, progress, if any, is even less visible. The current trend in template-free modeling is to identify fragments of known structures [29, 30] that might be locally similar to the target structure. Such fragments are subsequently used in fragment-recombination algorithms to build a full model of a protein. This approach tries to exploit the leverage of continuously growing knowledge of PDB structures, resulting to a large extent from the PSI projects.

Since we performed our analysis, PSI programs have transitioned to the PSI:Biography phase, focusing on proteins with biological and medical applications. As part of the new PSI Structural Biology Knowledgebase, a new project CAMEO (Continuous Automated Model EvaluatiOn) [31] is being developed in order to provide an on-going comparison of structure predictions for PSI targets by various servers. However, the full results of this study are not yet available. In light of the changes in the focus of SG, and the addition of new structure prediction servers, it will be very interesting to see if our findings will be confirmed by the CAMEO results.

Conclusions

In this study we surveyed fully automated template-based structure prediction for almost two hundred sequences and compared them to the experimentally determined structures. Our results render a general picture of the strengths and weaknesses of these services. In general, even using templates that are highly similar in sequence, the predicted model for a target will not be an accurate match to the known structure over the whole polypeptide sequence; in most cases there are structural regions that cannot be predicted accurately, e.g. flexible tails and loops. When modeling a single domain, however, the probability that a large majority of a structure can be accurately modeled depends strongly on whether a template with high sequence identity can be found. Virtually all predicted single-domain proteins had models with a GDT(2Å) value (to the known structure) between 0.9 and 1.0 when the sequence of the template was >30% identical to the target, and many models built with templates that were >25% identical to the target were that accurate.

The major limitation of the current automated template structure prediction process is domain parsing; the servers performed much better given single domain proteins as input. Modeling each domain separately would most likely result in detecting better templates, but this process is relatively complicated and difficult to automate. In a domain-aware approach, one should run domain prediction calculations and (based on the results) define the boundaries between domains. However, the results of these domain prediction calculations are often ambiguous and thus difficult to integrate into an automated pipeline. If the target can be successfully split into domains, each of the domain sequences should be submitted to prediction servers separately. This would change a single step procedure (metaserver submission), into a three step one: domain decomposition, metaserver submission, and domain spatial arrangement. We are not aware of a reliable automated system capable of this procedure.

More elaborate template-modeling software, such as CABS [8, 32] or ROSETTA (used in an automated fashion; [33]) could probably give slightly better results, as demonstrated by

the server participants in the CASP8 experiment. Better modeling algorithms can better adapt the template structure to the new (target) sequence and, to some extent, fix sequence alignment errors. However, proper domain parsing and template selection would still remain the primary factor limiting the accuracy of the predicted models.

In this experiment we have also addressed whether structure prediction could have been substituted for some of the workload of SG. What percentage of the SG output in the time frame considered could have been satisfactorily predicted beforehand, without performing the actual experimental structure determination? The answer depends on what is considered a satisfactory prediction. If all one is interested in is determination of the protein fold, models do not need to be of very high accuracy. However, in many cases biologists are not interested in the overall fold but rather particular regions of a protein (ligand binding sites, catalytically active residues, etc). Ultimately, experimental structure determination will always trump structure prediction if the experimental structure is available.

Furthermore, template-based modeling methods are dependent on the throughput of experimental structures from structural biology in order to be successful, resulting in a synergy between the two methods. On the one hand, template-based modeling fills in the gaps in structural characterization of the protein universe between the targets determined by SG, and thus helps SG by reducing the number of potential targets. On the other, a significant fraction of all macromolecular structures have been solved by SG, providing a significant number and variety of new templates. Indeed, in recent years SG has provided most of the targets for the CASP competition [6].

However, in a recent study, Levitt [34] determined that clustering chain sequences at the 25% sequence identity threshold was a very good determinant for classifying proteins in SCOP families [35]. We found that most of the targets above 25% reference sequence identity could be modeled to within $GDT(2\text{\AA}) > 0.7$ (Fig. 1D), suggesting that the protein fold for such targets could be satisfactorily predicted. Such models can be used (for example) to design site-directed mutants or for NMR structure refinement. Sometimes these models may be used for virtual screening and docking small ligands [36]. Our results also suggest that the 30% sequence identity threshold used by SG centers could be lowered to about 25%, which is in agreement with the long-known result that for sequences of adequate length (>80 amino acids), 25% sequence identity is a reasonable cutoff for a high probability of structural homology [20].

For targets below 20% reference sequence identity, the best $GDT(2\text{\AA})$ is below 0.6 (Fig. 1D), indicating that such targets (which constitute the main focus of the structural coverage programs of SG centers) cannot be reliably predicted with the template-based metaservers. The range between 20% and 30% reference sequence identity is the transition zone between very incomplete models on the low end, and relatively good predictions (sufficient for fold determination) on the high end. The need for structure accuracy is much greater for proteins of biomedical relevance constituting an important and growing fraction of SG targets. Such targets often have relatively high reference sequence identity to a template. In our study, none of the servers examined achieved consistent quality of prediction within this range of sequence identity.

Acknowledgments

The authors would like to thank Alex Wlodawer and Rachel Vigour for valuable comments on the manuscript. This work was supported by grants GM94585, GM74942 and GM53163.

Abbreviations

SG	Structural Genomics
PDB	Protein Data Bank
CASP	Critical Assessment of Protein Structure Prediction
PSI	Protein Structure Initiative
CRMSD	Coordinate Root Mean Square Deviation
GDT	Greatest Distance Test

References

- Grabowski M, et al. Structural genomics: keeping up with expanding knowledge of the protein universe. *Curr Opin Struct Biol.* 2007; 17(3):347–353. [PubMed: 17587562]
- Levitt M. Nature of the protein universe. *Proc Natl Acad Sci U S A.* 2009; 106(27):11079–11084. [PubMed: 19541617]
- Anfinsen CB. Principles that govern the folding of protein chains. *Science.* 1973; 181(96):223–230. [PubMed: 4124164]
- Bradley P, Misura KM, Baker D. Toward high-resolution de novo structure prediction for small proteins. *Science.* 2005; 309(5742):1868–1871. [PubMed: 16166519]
- Vitkup D, et al. Completeness in structural genomics. *Nat Struct Biol.* 2001; 8(6):559–566. [PubMed: 11373627]
- Moult J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struct Biol.* 2005
- Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993; 234(3):779–815. [PubMed: 8254673]
- Kolinski A, Gront D. Comparative modeling without implicit sequence alignments. *Bioinformatics.* 2007; 23(19):2522–2527. [PubMed: 17660201]
- Bujnicki JM, et al. Structure prediction meta server. *Bioinformatics.* 2001; 17(8):750–751. [PubMed: 11524381]
- Chen L, et al. TargetDB: a target registration database for structural genomics projects. *Bioinformatics.* 2004; 20(16):2860–2862. [PubMed: 15130928]
- Ginalski K, et al. 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics.* 2003; 19(8):1015–1018. [PubMed: 12761065]
- Kurowski MA, Bujnicki JM. GeneSilico protein structure prediction meta-server. *Nucleic Acids Res.* 2003; 31(13):3305–3307. [PubMed: 12824313]
- Wallner B, Elofsson A. Pcons5: combining consensus, structural evaluation and fold recognition scores. *Bioinformatics.* 2005; 21(23):4248–4254. [PubMed: 16204344]
- Alexandrov N, Shindyalov I. PDP: protein domain parser. *Bioinformatics.* 2003; 19(3):429–430. [PubMed: 12584135]
- Kabsch W. Solution for best rotation to relate two sets of vectors. *Acta Crystallogr. A.* 1976; 32:922–923.
- Gront D, Kolinski A. BioShell—a package of tools for structural biology computations. *Bioinformatics.* 2006; 22(5):621–622. [PubMed: 16407320]

17. Gront D, Kolinski A. Utility library for structural bioinformatics. *Bioinformatics*. 2008; 24(4):584–585. [PubMed: 18227118]
18. Smith TF, Waterman MS. Identification of common molecular subsequences. *Journal of molecular biology*. 1981; 147(1):195–197. [PubMed: 7265238]
19. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*. 1992; 89(22):10915–10919. [PubMed: 1438297]
20. Sander C, Schneider R. Database of homology-derived protein sequences and the structural meaning of sequence alignment. *Proteins: Structure, Function, and Genetics*. 1991; 9:56–68.
21. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*. 1997; 25(17):3389–3402. [PubMed: 9254694]
22. Moult J, et al. Critical assessment of methods of protein structure prediction-Round VIII. *Proteins-Structure Function and Bioinformatics*. 2009; 77:1–4.
23. Venclovas C, Margelevicius M. The use of automatic tools and human expertise in template-based modeling of CASP8 target proteins. *Proteins-Structure Function and Bioinformatics*. 2009; 77:81–88.
24. Moult J, et al. Critical assessment of methods of protein structure prediction (CASP)--round IX. *Proteins*. 2011; 79(Suppl 10):1–5. [PubMed: 21997831]
25. Kryzhafovych A, Fidelis K, Moult J. CASP9 results compared to those of previous CASP experiments. *Proteins*. 2011; 79(Suppl 10):196–207. [PubMed: 21997643]
26. Moult J, et al. Critical assessment of methods of protein structure prediction-Round VIII. *Proteins*. 2009; 77(Suppl 9):1–4. [PubMed: 19774620]
27. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics*. 2005; 21(7): 951–960. [PubMed: 15531603]
28. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008; 9:40. [PubMed: 18215316]
29. Xu D, Zhang Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins*. 2012; 80(7):1715–1735. [PubMed: 22411565]
30. Gront D, et al. Generalized fragment picking in Rosetta: design, protocols and applications. *PLoS One*. 2011; 6(8):e23294. [PubMed: 21887241]
31. Cameo Project. 2012 Available from: www.cameo3d.org.
32. Kolinski A. Protein modeling and structure prediction with a reduced representation. *Acta Biochim Pol*. 2004; 51(2):349–371. [PubMed: 15218533]
33. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*. 2004; 32(Web Server issue):W526–W531. [PubMed: 15215442]
34. Levitt M. Growth of novel protein structural data. *Proc Natl Acad Sci U S A*. 2007; 104(9):3183–3188. [PubMed: 17360626]
35. Murzin AG, et al. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*. 1995; 247(4):536–540. [PubMed: 7723011]
36. Baker D, Sali A. Protein structure prediction and structural genomics. *Science*. 2001; 294(5540): 93–96. [PubMed: 11588250]

Summary

We analyzed automatic structure prediction by template-based metasevers by comparing the resulting structures with the corresponding structures solved experimentally by structural genomics centers. Accuracy of the predicted model was satisfactory (70% of the main chain predicted within 2 Å) when template sequence identity relative to the target was greater than 25%. The majority of structural genomics targets, however, were not accurately predicted by the metasevers.

Figure 1.

Structural similarity of predicted models as a function of sequence identity. **(A)** All predicted models plotted by crmsd of predicted versus experimental model, as a function of alignment sequence identity (ASI; i.e. sequence identity between the input sequence and the sequence of the template structure). The most accurate models by crmsd for each input sequence are shown in blue, the others in red. **(B)** All predicted models plotted by global distance test (GDT) fraction with a 1 Å cutoff, as a function of ASI. The most accurate models by GDT(1Å) for each input sequence are shown in blue, the others in red. **(C)** All predicted models plotted by GDT with a 1 Å cutoff, as a function of reference sequence identity (RSI; i.e. sequence identity between the input sequence and the sequence of the best template structure). The most accurate models by GDT(1Å) for each input sequence are shown in blue, the models ranked as best for each input sequence by the metaservers are in green, and the other models are in red. **(D)** The same data as in Fig. 1C save that a GDT with a 2 Å cutoff is used.

Figure 2.

Modeling results for subset of PSI targets with multiple (3 or more) templates available. Each vertical set of bars represents models built using the three best templates (with at least 5 models built on each). The vertical ranking of the three bars was determined by the highest GDT(1Å) of the models built on each template. The blue bars represent the range of GDT(1Å) values for models built on the best template, green for the range for the second best template, and red bars for the third best template. The bars are connected by a thin red line. For some targets, only 2 templates had 5 or more models built on them. The targets are listed horizontally in order of increasing overall mean GDT(1Å) value.

Figure 3. Superposition of the experimental models for the PDB deposits 3CQB chains A and B with identical sequences. The crmsd distance between these models is $5.2 \text{ \AA}$, with $\text{GDT}(1\text{\AA}) = 0.70$ and $\text{GDT}(2\text{\AA}) = 0.83$. The discrepancy is due to unordered N-terminal regions in both models. This example illustrates the usefulness of GDT over crmsd.

Table 1

Modeling results for the targets with reference sequence identity greater than 30%.

center	Target				best template				Results			
	PDB code	chain	Xray/ NMR	released	PDB code	chain	Xray/ NMR	released	GDT(1)	GDT(2)	crmsd	ref. seq id
BSGI	2OZN	A	X	2008-05-06	2JH2	A	X	2007-11-06	0.917	0.985	0.67	1.000
BSGI	2OZN	B	X	2008-05-06	2JNK	A	N	2008-01-29	0.229	0.481	2.76	1.000
NESGC	3D3N	B	X	2008-06-10	3BXP	A	X	2008-01-29	0.866	0.906	2.68	0.992
NYSGRG	3DBI	A	X	2008-07-01	3BRQ	A	X	2008-01-22	0.482	0.704	2.71	0.975
JCSG	3D4O	C	X	2008-07-08	2RIR	A	X	2007-10-23	0.483	0.555	2.77	0.636
JCSG	3DO6	A	X	2008-09-02	1FPM	A	X	2001-08-31	0.677	0.919	1.31	0.540
NYSGRG	3CKW	A	X	2008-04-29	3COM	A	X	2008-04-15	0.635	0.863	1.48	0.518
NYSGRG	3DIP	A	X	2008-07-29	2POD	A	X	2007-06-05	0.733	0.898	1.97	0.466
JCSG	3D7Q	B	X	2008-06-17	2NLV	A	X	2006-12-05	0.748	0.929	1.35	0.455
NESGC	3DJB	A	X	2008-08-19	2QGS	A	X	2007-07-17	0.644	0.819	2.18	0.441
NESGC	3D3U	A	X	2008-07-15	2OAS	A	X	2007-01-23	0.501	0.918	1.72	0.439
NESGC	3DAI	A	X	2008-07-29	2RGO	A	X	2008-01-15	0.520	0.849	2.01	0.438
NESGC	3DIO	B	X	2008-09-09	2PIQ	A	X	2007-05-01	0.540	0.810	3.03	0.429
MCSG	3DCI	C	X	2008-09-30	2QOQ	A	X	2007-12-11	0.630	0.856	1.99	0.428
NYSGRG	3DFH	A	X	2008-07-01	3BSM	A	X	2008-01-15	0.642	0.962	1.30	0.409
NYSGRG	3DJC	B	X	2008-07-15	2H3G	X	X	2007-03-20	0.584	0.816	1.98	0.402
MCSG	3DED	C	X	2008-08-05	2PLS	A	X	2007-05-22	0.549	0.747	1.94	0.385
JCSG	3DO5	A	X	2008-09-02	2EJW	A	X	2008-02-12	0.419	0.653	2.82	0.378
NESGC	3D3Q	A	X	2008-07-15	2QGN	A	X	2007-07-17	0.437	0.679	1.88	0.378
NYSGRG	3CZB	B	X	2008-06-10	2G5D	A	X	2006-03-21	0.267	0.464	5.20	0.374
NESGC	3DL3	F	X	2008-08-26	3BB6	A	X	2007-11-20	0.664	0.864	1.57	0.373
JCSG	3DUK	D	X	2008-12-23	3BLZ	A	X	2007-12-25	0.750	0.871	1.89	0.370
NESGC	2K3I	A	N	2008-07-08	2JZ5	A	N	2008-02-19	0.444	0.596	11.42	0.367
NYSGRG	3CPG	A	X	2008-05-13	1W8G	A	X	2006-07-05	0.447	0.722	2.02	0.356
JCSG	3D8P	B	X	2008-10-28	2Q7B	A	X	2007-06-26	0.594	0.850	2.08	0.350

center	Target			best template					Results			
	PDB code	chain	Xray/ NMR	released	PDB code	chain	Xray/ NMR	released	GDT(1)	GDT(2)	crmsd	ref. seq id
NYSGRC	3DOU	A	X	2008-09-02	1EIZ	A	X	2000-08-30	0.629	0.840	1.76	0.349
MCSG	3CTV	A	X	2008-04-29	3HDH	A	X	1999-10-08	0.500	0.772	3.04	0.348
NESGC	2K57	A	N	2008-09-30	3BDU	A	X	2007-11-27	0.755	0.962	0.86	0.345
MCSG	3DDH	A	X	2008-09-30	2PKE	A	X	2007-05-08	0.425	0.768	2.60	0.341
NYSGRC	3DLS	A	X	2008-08-26	2HAK	A	X	2006-07-11	0.358	0.642	7.19	0.340
JCSG	3DFE	B	X	2008-08-05	2CZ4	A	X	2006-01-10	0.687	0.940	1.29	0.337
NYSGRC	3CYJ	D	X	2008-05-06	2OVL	A	X	2007-03-20	0.955	0.981	0.58	0.337
NESGC	3CWI	A	X	2008-05-06	2CU3	A	X	2006-05-23	0.667	0.870	1.85	0.333
MCSG	3D7N	A	X	2008-07-15	1ZWK	A	X	2005-06-14	0.688	0.948	3.03	0.331
NESGC	3CWQ	A	X	2008-05-13	1WCV	I	X	2005-01-12	0.388	0.567	3.74	0.330
JCSG	3DCF	A	X	2008-07-15	3CCY	A	X	2008-03-18	0.278	0.605	3.75	0.330
MCSG	3D3R	A	X	2008-05-27	2Z1C	A	X	2007-07-17	0.395	0.500	3.56	0.325
MCSG	3D01	L	X	2008-07-01	2OTM	A	X	2007-02-20	0.706	0.901	1.70	0.325
JCSG	3D82	E	X	2008-06-10	2I45	A	X	2006-09-19	0.594	0.842	1.38	0.324
JCSG	3DB0	B	X	2008-07-29	2FG9	A	X	2006-01-10	0.556	0.766	2.50	0.323
JCSG	3DMB	A	X	2008-08-26	2QEA	A	X	2007-07-10	0.397	0.628	3.07	0.322
NYSGRC	3CWV	A	X	2008-05-06	1KIJ	A	X	2002-06-03	0.418	0.693	3.74	0.318
NYSGRC	3DDM	A	X	2008-06-24	2OG9	A	X	2007-02-27	0.364	0.713	2.65	0.317
NYSGRC	3CTP	A	X	2008-05-06	1JFT	A	X	2002-02-08	0.522	0.901	2.50	0.316
NESGC	2K5N	A	N	2008-08-19	1MJC	A	X	1994-06-22	0.742	0.864	1.56	0.311
JCSG	3D2L	C	X	2008-05-20	1Y8C	A	X	2004-12-28	0.405	0.694	2.89	0.310
NESGC	3DM4	A	X	2008-08-26	1TXY	A	X	2004-11-16	0.663	0.857	1.98	0.306
NYSGRC	3CZ5	C	X	2008-05-06	1A04	A	X	1998-03-18	0.521	0.732	1.44	0.303
NESGC	3D0W	A	X	2008-05-20	1NWM	X	X	2003-03-25	0.261	0.326	6.95	0.302
NESGC	2K54	A	N	2008-08-19	1OGZ	A	X	2003-09-04	0.496	0.740	2.32	0.301
MCSG	3D1P	A	X	2008-07-08	1CWR	A	X	2000-08-28	0.509	0.797	3.35	0.300

Table 2

The number of targets for which each server was able to find either the most accurate model or any model at all. Different servers used different template proteins to build models. Columns 2 and 3 show for how many targets (out of all 199 targets) a given server returned either the most accurate model or at least one model at all, respectively. Columns 4 and 5 show for how many of the 16 targets which exceed 40% reference sequence identity a given server returned either the best template model or at least one model at all, respectively.

Server	for all targets		for targets over 40% seq. ID.	
	best-template models	all models	best-template models	All models
3DPS	13	52	0	3
BASIC	134	194	13	16
Easy Pred	33	72	9	12
FFAS 3	127	193	13	16
FUGU	55	136	7	13
HHPred 2	38	118	5	11
INGBU	23	48	0	0
Meta-Genthrader	99	181	12	15
PDB-PsiBlast	98	161	13	16
RPSB	80	133	11	13
SAM6	14	26	2	2
ST02	68	141	10	15