PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

PREDICTION REPORT

The use of automatic tools and human expertise in template-based modeling of CASP8 target proteins

Česlovas Venclovas* and Mindaugas Margelevičius

Institute of Biotechnology, Graičiūno 8, LT-02241 Vilnius, Lithuania

ABSTRACT

Here, we describe our template-based protein modeling approach and its performance during the eighth community-wide experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP8, http://predictioncenter.org/casp8). In CASP8, our modeling approach was supplemented by the newly developed distant homology detection method based on sequence profile–profile comparison. Detection of structural homologs that could be used as modeling templates was largely achieved by automated profile-based searches. However, the other two major steps in template-based modeling (TBM) (selection of the best template(s) and construction of the optimal sequence-structure alignment) to a large degree relied on the combination of automatic tools and manual input. The analysis of 64 domains categorized by CASP8 assessors as TBM domains revealed that we missed correct structural templates for only four of them. The use of multiple templates or their fragments enabled us to improve over the structure of the single best PDB template in about 1/3 of our models for TBM domains. Our results for sequence-structure alignments are mixed. Although many models have optimal or near optimal sequence mapping, a large fraction contains one or more misaligned regions. Strikingly, in spite of this, our TBM models have the best overall alignment accuracy scores. This clearly suggests that the correct mapping of protein sequence onto three-dimensional structure remains one of the big challenges in protein structure prediction.

Proteins 2009; 00:000–000.

Key words: protein structure prediction; comparative modeling; profile–profile comparison; structural templates; sequence-structure alignment; 3D model; model evaluation; distant homology; alignment errors.

INTRODUCTION

The knowledge of the three-dimensional (3D) protein structure is important for understanding protein function, interactions, interpretation of experimental data, knowledge-based drug design, and in many other cases. Structural genomics centers with their industry-like approach towards structure determination have greatly contributed to a more dense coverage of protein sequence universe with structural representatives. However, even though the number of determined protein structures is steadily increasing, it still comprises a tiny fraction of all known protein sequences. It is obvious that, given present limitations of experimental structural characterization of proteins, only fraction of them will ever be studied in detail experimentally. In contrast, computational methods offer essentially unlimited potential for protein structural studies.

Because 3D structure is the most evolutionary conserved protein feature, it can serve as a template to produce a structural model of a related protein. At present, template-based or comparative modeling is the most accurate protein structure prediction method,[1,2] but it can be applied only if it is possible to detect relationship between the sequence of interest (target) and known structures. When the relationship is close, the detection of related templates is trivial. However, as evo-

The authors state no conflict of interest.
Grant sponsors: Lithuanian State Science and Studies Foundation, Howard Hughes Medical Institute.
*Correspondence to: Česlovas Venclovas, Institute of Biotechnology, Graičiūno 8, LT-02241 Vilnius, Lithuania. E-mail: venclovas@ibt.lt
Received 5 April 2009; Revised 9 June 2009; Accepted 14 June 2009
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22515

lutionary distance increases, the detection of related templates becomes a limiting factor for the applicability of comparative modeling. Provided that distantly related structures are reliably detected, the quality of the template-based model (TBM) is primarily determined by the accuracy of the alignment between the target sequence and the structural template(s). The selection of the optimal structural template (or a set of templates) is also a significant factor determining model quality, especially in the case of remote relationships.

In previous CASPs, we have been mainly focusing on the two issues critical for the TBM, namely on the accuracy of sequence-structure alignments and the optimal template selection,[3–6] essentially leaving out the problem of the template detection. Lately, we have been applying our efforts to address this problem as well. In particular, we have been working on a new profile–profile comparison method featuring a number of novel theoretical and algorithmic developments (manuscript in preparation). Availability of an operational initial version of the new method (COMA; Comparison Of Multiple Alignments) before the start of the CASP8 experiment prompted us to test how the remote template detection by COMA can expand our capabilities in the TBM. In addition, we viewed CASP8 as an excellent opportunity to obtain hints as to what further improvements in COMA are most needed. Last but not least, we were interested to find out whether human expert input is beneficial and in which cases.

METHODS

In CASP8, we modeled target proteins in both automatic server and human expert modes. In the automatic mode, we tested our new method, implemented as two different servers (COMA and COMA-M). Because of the very short time frame available to us, both servers were set up as fairly simple tools for converting sequence-structure alignments into 3D models using MODELLER[7] for automatic model construction and Prosa2003[8] for model ranking. The main difference between the two is in handling the structural templates. COMA-M was set up to be able to use multiple templates, whereas models produced by COMA were always based on a single template.

In this report, we focus on our human expert mode approach (predictor group "IBT_LT"), which is summarized as a flowchart in Figure 1. In this approach, the type and the extent of human input varied depending on the target difficulty and the results of the assessment of COMA/COMA-M models in the context of other CASP8 servers. The simplest scenario (Fig. 1, left) includes a few cases when closely related structures were available and sequence-structure alignments were trivial. If our server models in those cases had no obvious flaws and fared well relative to those obtained by other automatic methods, little or no human intervention was used. Another scenario (Fig. 1, right) comprises targets, for which

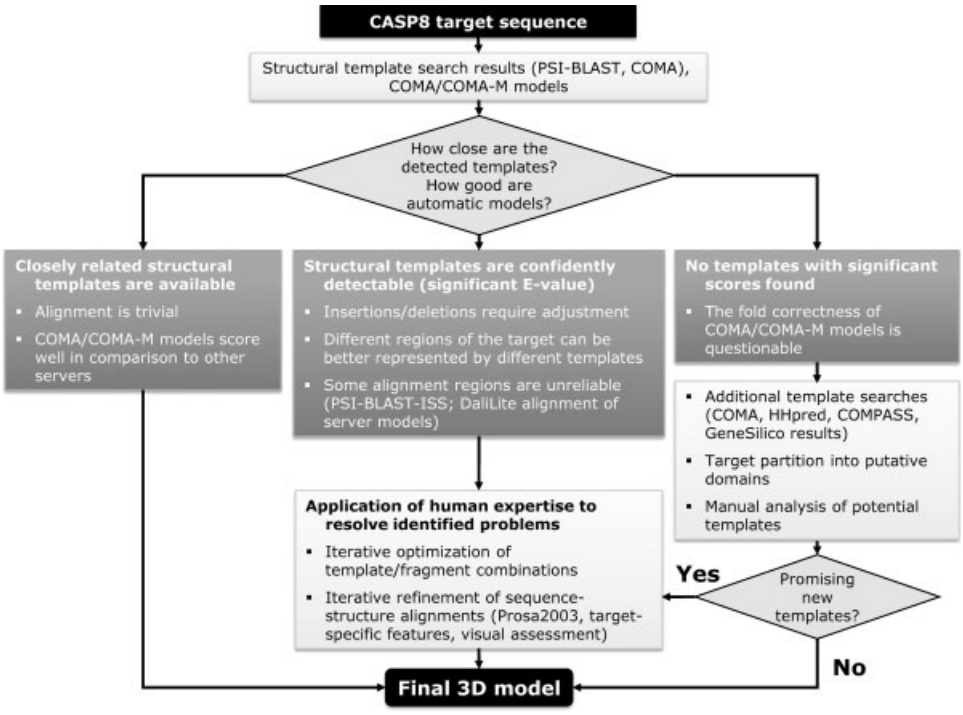

Figure 1

A flowchart of the human expert approach for predicting CASP8 targets.

COMA could not reliably detect any structural homolog (template), suggesting that our automatically generated server models may well have been completely wrong. In those cases, additional COMA searches were performed, including different variants of input multiple sequence alignments and/or search parameters. A common strategy was also to check HHpred,[9] COMPASS,[10] and the GeneSilico[11] metaserver results. If all these additional steps did not help to find any reliable structural match, then simply one of the unreliable COMA-M models having the best score was submitted to the Prediction Center. However, by far the largest pool of CASP8 targets consisted of those for which structural templates could be readily detectable by either PSI-BLAST[12] or COMA, but corresponding sequence-structure alignments were uncertain in one or more regions and/or there was a need to make a selection from available templates. Models for those targets were constructed independently using a manual procedure described below. In those cases, CASP8 server models including ours were used to identify problematic regions of the sequence-structure alignment and to provide a baseline in the evaluation of IBT_LT models.

Template detection and selection

If structural homologs (templates) could be easily found with PSI-BLAST, their selection was usually based on the consensus results of transitive searches carried out using the PSI-BLAST-ISS tool.[13] In other words, the most frequently detected structure during these transitive searches was considered to be the best template. If multiple templates were available, up to four structures that would introduce sufficient conformational variability were selected. However, if one of them appeared to be much closer to the target than the others, only this single template was used. When PSI-BLAST-ISS failed to detect any templates with significant *E*-values, those identified reliably by COMA were considered. In cases of multiple templates, the selection was an iterative process and the final set of templates were optimized based on the evaluation of corresponding models (section Model Evaluation). In some cases, fragments of either templates or server models were used, in addition to intact templates, to better represent a particular region of the model.

Sequence-structure alignments

This section describes the procedure, which was used to generate sequence-structure alignments when we were confident that we had detected at least one suitable template. Unless the alignment was trivial, reliably aligned regions were first identified with PSI-BLAST-ISS. In parallel, automatic server models downloaded from the CASP8 web site (http://predictioncenter.org/casp8) were superimposed with one of the representative templates using DaliLite,[14] and all the corresponding pairwise alignments were merged into a single, PSI-BLAST-ISS-like alignment. In such an alignment, the template sequence is aligned with multiple instances of the target sequence according to respective DaliLite structure-based template-model alignments. Like in PSI-BLAST-ISS, a good agreement between different models was considered to be an indicator of a reliable sequence-structure alignment region. For an alignment region to be treated as reliable, it was sufficient that at least one of the two aforementioned methods would produce a good consensus. In most cases, it was the method based on structure-based alignment of CASP8 server models that resulted in assignment of more extensive reliable regions. For the remaining (unreliably aligned) regions, alternative alignment variants were evaluated at the level of 3D models and the best variant retained in the final model. Not in every case the evaluation of alternative models for the target was showing a clear preference toward a particular alignment variant. To resolve the ambiguity in situations like this, a number of target homologs (no more than 10) were then selected, such that their alignment with the target would be well defined. The same alignment variants as tested for the target sequence were then used to construct models for the homologs. The best alignment variant was then picked according to the consensus results of evaluation of corresponding models.

Model construction

Three-dimensional structures were constructed automatically from sequence-template(s) alignments using the standard MODELLER protocol. Residue side chains were positioned with SCWRL3.[15] No further model optimization was performed.

Model evaluation

Model assessment was a central aspect of our human expert approach. The most important role of the model evaluation was in choosing the best alignment variant from a number of alternatives in uncertain regions. Model evaluation was also used to pick the best template or the best set of multiple templates. For both alignment and template assessment, alternative choices were evaluated by first constructing corresponding models and then assessing their energies and structural properties. The overall quality of a model was estimated by calculating its Prosa2003 energy *Z*-score and comparing it to *Z*-scores of the modeling template(s). Prosa2003 *Z*-scores were also used to assess how our models compare to corresponding models of automatic CASP8 servers. The targeted goal was to either match or improve over the *Z*-score of the most favorably assessed server model. Prosa2003 position-dependent energy profiles were used to detect local

flaws in modeled structures and thus guide the visual inspection. The manual analysis was an important component of model evaluation, especially in finalizing sequence-structure alignments. The benefit of manual analysis is in its ability to simultaneously assess those multiple structural features that may not be easily captured by the plain energy estimation. Optimization of both the set of templates and the sequence-structure alignment was performed in an iterative manner until model scores could not be significantly improved and the final model looked acceptable by the visual analysis.

RESULTS AND DISCUSSION

We have made predictions for all 57 CASP8 target proteins assigned for the human expert track, but some of these target proteins were cancelled by the CASP8 organizers and assessors. As the evaluation of 3D models in CASP is done at the level of individual structural domains, multidomain proteins usually have more than one "evaluation unit." Our predictions covered 70 of the final set of 71 evaluation units in the human expert track. The missing prediction was for the first domain of T0397 (T0397-D1). This is one of only seven domains classified by CASP8 assessors as free-modeling (FM) targets, for which no obvious similar structural templates could be identified after the target structures (answers) became available. The majority (64 domains) were classified as TBM targets, indicating that at least a distantly related structural template was available in the PDB at the time of the CASP8 prediction season.

Because all our models were constructed by explicitly utilizing structural templates, it is no surprise that our predictions for FM domains are quite poor. Prediction of only one FM domain (T0465-D1) received the GDT_TS[16] score better than average, the remaining five are worse. In our predictions for FM domains, only short fragments at best display a detectable similarity to the target structures.

Therefore, in this report, we focus only on our predictions for 64 CASP8 TBM domains. The overall results are summarized in Figure 2. Histograms show both the absolute quality of our models according to the GDT_TS score and the comparison to the overall top models that may be considered to represent the current state-of-art in protein structure prediction (here and throughout the article our analysis is based only on the most confident models (model 1)). Our results for TBM domains are in sharp contrast to FM domain predictions. Almost two-thirds of our models are either the best or closely approach the best ones (within 5% of GDT_TS). Twenty-one of our models appear among the overall best three, and they are most densely clustered within the more difficult TBM domains. This correlates with our major interest in exploiting distant evolutionary relationships in our research outside CASP. Among TBM domains predicted least successfully relative to the best CASP8 results, the notable exceptions are T0460-D1, T0466-D1, T0468-D1 and T0496-D2, each more than 20% of GDT_TS away from the top models.

Unfortunately, Figure 2 does not tell us anything about the relative contribution of the template choice and the

Figure 2

Performance by the IBT_LT group on 64 TBM domains according to the GDT_TS scores. Cyan and green bars denote GDT_TS values for each TBM domain; green color indicates that IBT_LT model appears among the three best ones. The orange color indicates by how much the overall best model outscores the IBT_LT model for the corresponding domain.

alignment accuracy, the two key factors in the TBM, to the success or failure. It also does not tell whether our models that appear good in the context of CASP8 results are approaching the limits defined by the closest template(s) or there still is a lot of room for improvement. To explore the impact of each factor separately, we decided to construct two additional models for each TBM domain. The idea behind the two additional models was to have the same metric, the GDT_TS score, for measuring the effect of both factors.

To estimate the alignment factor, for each TBM domain we constructed a model using the accepted IBT_LT CASP8 model as the template and the alignment that was obtained from DaliLite target-model superposition. The rationale for producing this type of model is that the structure of the model is not expected to change much, yet alignment may be "adjusted" according to the target-model structure comparison. If the original model had alignment errors, the newly obtained model based on the "adjusted" alignment would be expected to have a better GDT_TS score.

The purpose of constructing the second type of models was to find out what the model quality would have been, provided the structurally closest template was picked and optimally aligned. For this, each TBM domain was searched with DaliLite against the PDB (filtered at 90% sequence similarity).[17] Up to 10 structures (available at the actual prediction time frame) with best Dali *Z*-scores were then used in turn to construct models for a given domain based on the structure-based alignment. Model with the highest GDT_TS value was considered to represent the best template choice with the optimal sequence-structure alignment.

Comparison of the GDT_TS values of the original CASP8 models and the corresponding ones remodeled according to the "adjusted" DaliLite alignment indicates whether the alignment could have been improved or not. Likewise, contrasting scores of models based on "adjusted" alignments with those representing the best template choice show whether the template selection was good or bad. Combined data are plotted in Figure 3, in which the vertical axis is the estimation of the alignment quality, whereas the horizontal axis provides an estimate of how effective was the template selection.

Negative values on the vertical axis indicate that models based on "adjusted" alignments score higher according to GDT_TS, suggesting that it was possible to make CASP8 models more accurate by improving alignment alone. Positive values on the same axis indicate that sometimes a structure-based alignment may make a model slightly worse. In this context, it implies that the alignment used for modeling could not have been improved. Negative values on the horizontal axis indicate that the model could have been more accurate if the best available template was used, whereas positive values mean that the CASP8 model is an improvement over the best template.

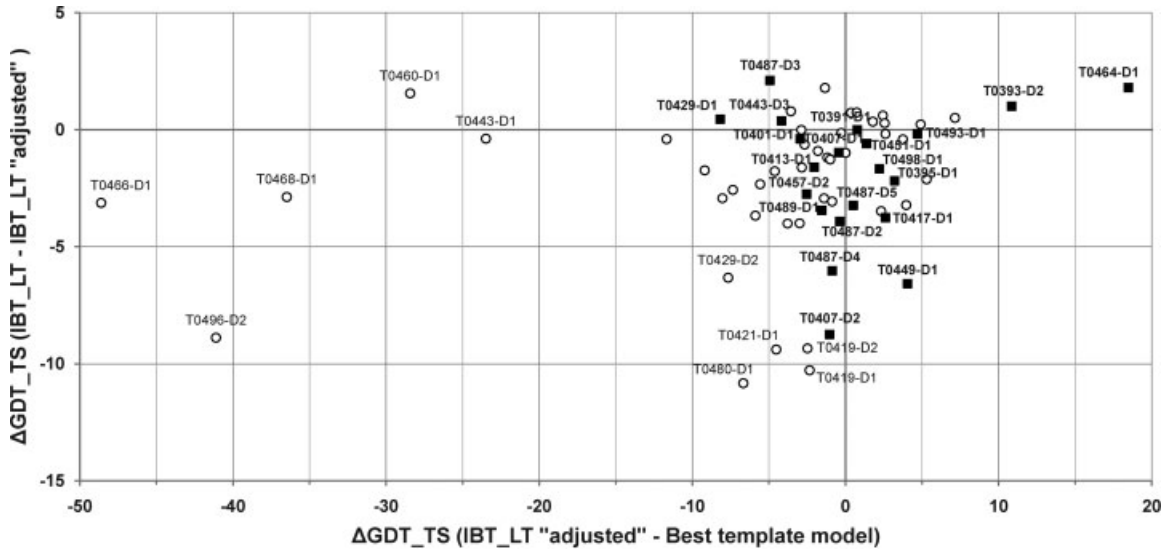

Figure 3

Estimation of the alignment correctness and the efficiency of picking good templates. Vertical axis represents the difference in GDT_TS between the deposited CASP8 models and corresponding models after the structure-based "adjustment" of sequence mapping. Horizontal axis shows the difference in GDT_TS between models derived from the "adjusted" alignment and those obtained from the best structural templates. Filled squares and domain IDs in bold denote our models that are among the overall best three. Models that are the worst according to either template detection or alignment also display their IDs.

Template detection

Figure 3 shows that there are several outliers in the template quality. The four (T0466-D1, T0496-D2, T0468-D1, and T0460-D1) with most negative values are those that also score worst in comparison to the corresponding overall best models (see Fig. 2). Upon closer inspection, it turned out that these were the only TBM domains, for which our models were based on incorrect structural templates. T0460, T0466, and T0468 are all single-domain proteins that have related structures in the PDB. Why did we fail to find these relatives? T0460 (protein PF0246 from *Pyrococcus furiosus*) is a sequence singleton. Without a sequence profile COMA's sensitivity is dramatically reduced, so this distant relationship had little chance of being discovered. In contrast, for both T0466 and T0468, sequence profiles could be generated, but their relationship to OB-fold proteins was also missed. Interestingly, neither HHpred nor COMPASS, two other powerful profile-based methods, could find any reliable match for any of these three proteins, suggesting that for sequence-based methods these were very challenging targets. T0496-D2 is a short coiled-coil motif classified in Pfam as IDEAL due to the characteristic sequence motif. Although COMA did correctly recognize the α-helical nature of the C-terminal region, the actual selected motif had a different geometry. Another outlier, T0443-D1, having a better score, but still far from the optimal template, had been assigned partially correct structural motif, matching two of the three helices. All other TBM domain models are based on correct templates. Although many models still do not match optimal templates, a significant number of models show improvement over the best template. The two most successful cases are T0464 and T0393-D2, scoring over 10% of GDT_TS higher than corresponding optimal TBMs.

In many cases, our human expert group outperformed our own automatic servers. Is it because human input helped to find distant relationships? Surprisingly, we find that only in one case (T0443-D3) a relatively trivial human input, associated with the domain boundary recognition, was helpful to detect a distant template. For T0443, we noticed that there are distant homologs (annotated as "Coenzyme PQQ synthesis protein D") that are much shorter and match only to the C-terminal region of the T0443. Substitution of the query with one of these short proteins enabled COMA to produce a statistically significant match to the winged helix-turn-helix structural motif, a correct structural template. In another case, COMA servers did not submit any model for the C-terminal domain of T0407 (T0407-D2). Apparently, this failure was due to the simplicity of our server setup, which initially lacked any kind of protocol for partitioning the target sequence into putative domains. Once the sequence partitioning was implemented well into the CASP8, but before the T0407 expiration for human expert groups, the servers could correctly assign the immunoglobulin fold for the T0407 C-terminal region in a standard run. Thus, the superiority of human expert over automatic modeling mode was mainly in other modeling phases, such as more efficient selection/combination of detected templates and more accurate alignments.

Sequence-structure alignments

In CASP8, just like in previous CASPs, the sequence-structure alignment was our major emphasis. This is a very important issue if models are to be used to guide experiments at the level of individual residues, such as the site-directed mutagenesis. Analysis of our results shows that the consensus approach (PSI-BLAST-ISS or structure-based alignment of CASP8 server models) works well in defining reliable alignment regions that subsequently translate into well-predicted structural regions. Even such a simple consensus approach enabled us to avoid occasional alignment errors present in a fraction of the server models. However, a much more challenging task was to resolve the uncertainty regarding the alignment in unreliable regions. The plot in Figure 3 indicates that our alignments are optimal or nearly optimal for a number of TBM domains, but quite a few still have alignment errors. Upon closer inspection, the absolute majority of these erroneous regions fit the pattern observed in earlier CASP experiments (e.g., CASP5[5]). In particular, many of the alignment errors affect β-strands that are at the edges of β-sheets. Appearance or disappearance of β-bulges in respect to the template structure is another common cause of alignment errors. Misaligned helices are often highly solvent-exposed and/or shifted relative to the template. Apparently, alternative alignments in these error-prone regions have subtle energy differences and therefore are difficult to rank correctly.

In our view, the value of human expertise in the alignment step often is in the ability to exploit case-specific features that are not necessarily frequently seen or considered to be important for selecting the optimal alignment. In Figure 4, we present one such example, the alignment of the edge β-strand (β1) in T0413. According to the alignment accuracy data available at the Prediction Center, less than 15% groups mapped the sequence correctly onto this β-strand, indicating that this has been a challenging case. In retrospect, the N-terminal region of T0413, including approximate location of β1, displayed no consensus alignment either using PSI-BLAST-ISS or structure-based comparison of server models [Fig. 4(A)]. On the other hand, comparison of related structures revealed a structurally well-conserved central β-sheet suggesting that T0413 is not an exception. As in all cases, we applied energy assessment with Prosa2003 *Z*-scores and energy profiles, but that has not produced any more clarity as to the optimal sequence mapping onto the β-strand. However, in the unambiguous alignment of T0413 with closely related sequences, we noticed the

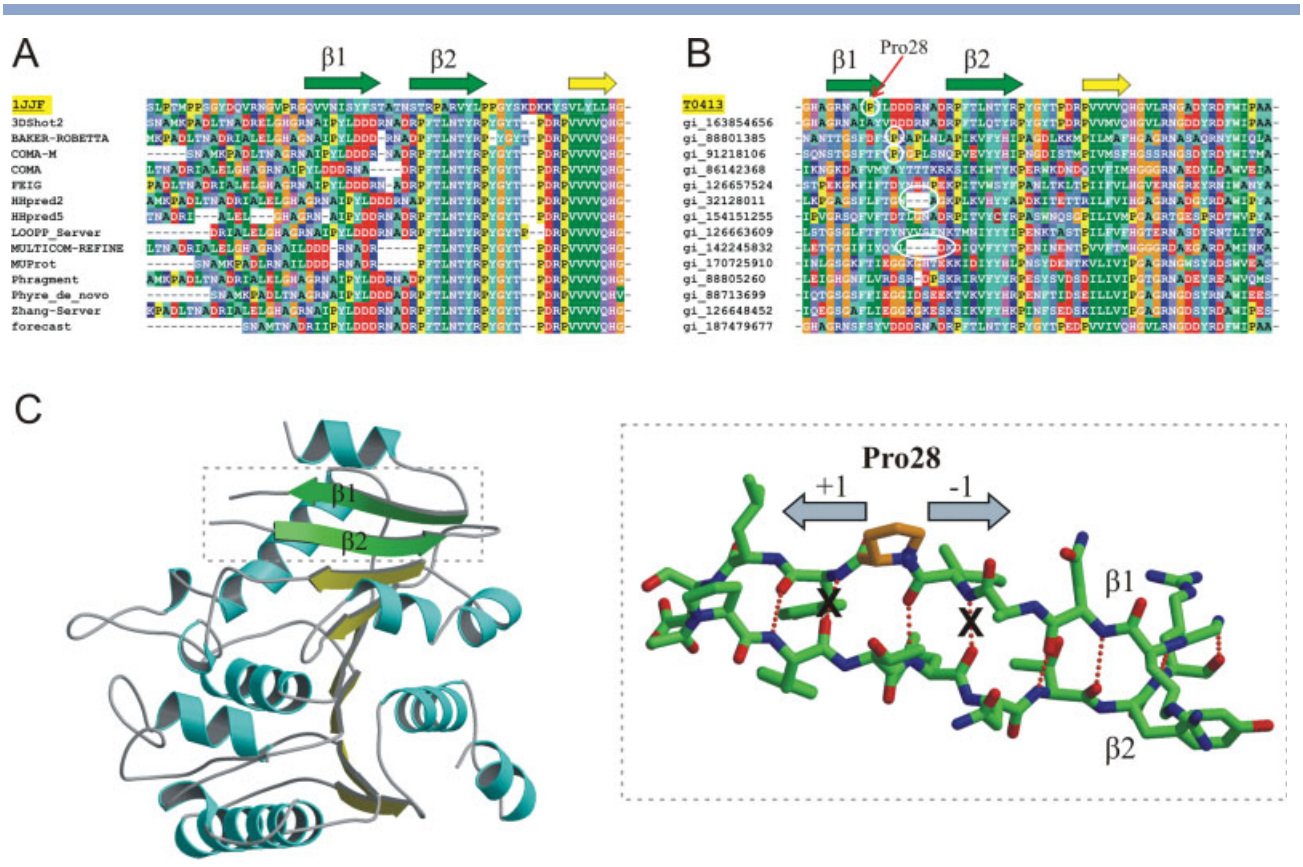

Figure 4

Resolving sequence-structure alignment uncertainty for the edge β-strand in T0413. (**A**) Merged structure-based alignments between one of the templates and a set of server models. Secondary structure of the template is mapped above the alignment. (**B**) Multiple sequence alignment of T0413 and its close homologs. Positions of Pro residues and gaps relevant to the alignment selection are indicated with white circles. Red arrow points to Pro28 in T0413. (**C**) T0413 3D structure. The first two β-strands are colored green and are shown in detailed representation on the right side. Black crosses indicate hydrogen bonds that would be lost upon Pro28 shift by one position in corresponding direction.

presence of Pro residues both in the target and some homologs within the region of the putative β1-strand [Fig. 4(B)]. Within the edge β-strand, the Pro residue can be accommodated without interfering with main chain hydrogen bonding only at every other position [Fig. 4(C)]. Thus, thanks to the guidance by Pro positions in the target and homologs and also by the positions of gaps in the alignment, it was a fairly straightforward task to find the unique alignment variant, which could accommodate Pro without disrupting hydrogen bonding pattern. This example gives at least a flavor of how human input has been helping us to make an informed selection from different alignment variants.

As we have been constructing sequence-structure alignments without relying on any particular automatic server, this "independent" mode was also helpful in understanding some of the causes of alignment errors made by COMA servers. In particular, the analysis of CASP8 results helped us to fix the problem associated with the comparison of two profiles derived from "thin" multiple sequence alignments (alignments that consist of only few aligned sequences and carry little evolutionary information).

What is more difficult: template detection or sequence-structure mapping?

It is interesting to consider the overall best models for TBM domains. Apparently, best models for all TBM domains are assigned the correct structural fold, independently whether this was achieved by finding a related template or assembling the structure without the use of explicit templates. However, many of those best models still have regions, in which residues are out of register. It would seem that matching a sequence with a specific 3D shape out of all possible shapes should be more difficult than finding a locally optimal mapping of residues. However, overall CASP results as well as our own experience suggest it is the opposite. Therefore, significant improvement in sequence-structure mapping would be among the most important breakthroughs in protein structure modeling.

CONCLUSIONS

We used CASP8 as testing grounds to find out whether there is a place for human expertise in protein structure prediction in comparison to currently best automatic methods. We find that human input may be the most valuable in the sequence-structure alignment step (e.g., T0413), because the correct mapping of a sequence onto the 3D structure seems to be a continuing serious problem. This problem seems to manifest itself regardless of whether structure prediction methods explicitly use templates or not. Most sequence mapping errors are often associated with structural elements that are peripheral, highly exposed to the solvent, and lack sequence conservation. Clearly, a more sensitive energy estimation methods coupled with the efficient conformational sampling are needed to significantly reduce the sequence-structure alignment problem. We found that human input was beneficial in one other aspect of protein modeling, an effective use of multiple templates. It led to an improvement over the single best template for a significant fraction of models. Considering the increasing numbers of PDB structures, this seems to be a promising way of making a "cheap" initial model "refinement." Human input seems to be of little use in two extremes: (1) high accuracy modeling and (2) those cases when no related structural template could be detected.

We also used CASP8 to test our own profile-based homology detection method, COMA. Results turned out to be very informative. The method appears to be very effective in distant homology detection. At the same time, CASP8 results helped us uncover some of its weak points related to the alignment quality in certain situations and make subsequent improvements.

ACKNOWLEDGMENTS

The authors thank CASP8 organizers, assessors, and experimental structural biologists for making this experiment possible. They also thank Ana Venclovienė and Albertas Timinskas for useful comments.

REFERENCES

1. Baker D, Sali A. Protein structure prediction and structural genomics. Science 2001;294:93–96.
2. Kryshtafovych A, Venclovas Č, Fidelis K, Moult J. Progress over the first decade of CASP experiments. Proteins 2005;61(Suppl 7):225–236.
3. Venclovas Č, Ginalski K, Fidelis K. Addressing the issue of sequence-to-structure alignments in comparative modeling of CASP3 target proteins. Proteins 1999;(Suppl 3):73–80.
4. Venclovas Č. Comparative modeling of CASP4 target proteins: combining results of sequence search with three-dimensional structure assessment. Proteins 2001;(Suppl 5):47–54.
5. Venclovas Č. Comparative modeling in CASP5: progress is evident, but alignment errors remain a significant hindrance. Proteins 2003;53(Suppl 6):380–388.
6. Venclovas Č, Margelevičius M. Comparative modeling in CASP6 using consensus approach to template selection, sequence-structure alignment, and structure assessment. Proteins 2005;61(Suppl 7):99–105.
7. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
8. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins 1993;17:355–362.
9. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 2005;33:W244–W248.
10. Sadreyev RI, Tang M, Kim BH, Grishin NV. COMPASS server for remote homology inference. Nucleic Acids Res 2007;35:W653–W658.
11. Kurowski MA, Bujnicki JM. GeneSilico protein structure prediction meta-server. Nucleic Acids Res 2003;31:3305–3307.
12. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
13. Margelevičius M, Venclovas Č. PSI-BLAST-ISS: an intermediate sequence search tool for estimation of the position-specific alignment reliability. BMC Bioinform 2005;6:185.
14. Holm L, Park J. DaliLite workbench for protein structure comparison. Bioinformatics 2000;16:566–567.
15. Canutescu AA, Shelenkov AA, Dunbrack RL, Jr. A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci 2003;12:2001–2014.
16. Zemla A, Venclovas Č, Moult J, Fidelis K. Processing and analysis of CASP3 protein structure predictions. Proteins 1999;(Suppl 3):22–29.
17. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v. 3. Bioinformatics 2008;24:2780–2781.