PROTEINS: Structure, Function, and Genetics Suppl 5:47–54 (2001)
DOI 10.1002/prot.10008

Comparative Modeling of CASP4 Target Proteins: Combining Results of Sequence Search With Three-Dimensional Structure Assessment

Česlovas Venclovas*
Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California

ABSTRACT **Comparative modeling aims at constructing molecular models for proteins of unknown structure, by using known structures of related proteins as templates. To test the comparative modeling approach reported here, predictions for 13 target proteins were submitted during the fourth round of "blind" protein structure prediction experiment (CASP4; http://PredictionCenter.llnl.gov/casp4). Sequence identity between these target proteins and the closest known structures ranged from 13 to 58%, indicating a broad spectrum of prediction difficulty. Although this broad difficulty range required addressing a variety of issues, the most important proved to be sequence-structure alignment for distant homology targets. The alignment step was based on structure-based evaluation of alignment variants produced mainly with PSI-BLAST intermediate sequence search procedure (PSI-BLAST-ISS). Although a fraction of correctly aligned residues in resulting models was markedly better than the average in all cases, for distant homology targets it was still considerably below the estimated achievable level. Results with CASP4 targets show that, along with the correctness of sequence-structure alignments, effective use of multiple template structures may significantly increase accuracy of the model structure. Improvement in this area should also result in more accurate loop modeling and side-chain prediction. Proteins 2001; Suppl 5:47–54.** © 2002 Wiley-Liss, Inc.

Key words: protein structure prediction; sequence-structure alignment; 3D model; model evaluation; distant homology; alignment errors

INTRODUCTION

Comparative modeling of three-dimensional (3D) protein structures is based on the observation that related protein sequences adopt the same general fold. Once the experimental structure is determined for at least one representative of a protein sequence family, structures for other related proteins could be then modeled by comparison. With protein sequences pouring as a result of genome sequencing projects, but with experimental structure determination lagging far behind, comparative modeling becomes a method of choice to structurally characterize many new protein sequences. Generally speaking, there are no inherent limits as to the application of comparative modeling. In practice, effective limits are imposed by the ability to detect relatedness between query protein sequence (target) and any of the proteins with known 3D structure (templates). The issues that have to be addressed in comparative modeling to a large degree depend on how closely related are the target and the templates. In high homology cases, target backbone structure is usually expected to be very similar to that of the template, so that positioning residue side chains and modeling few (if any) insertions or deletions is the major emphasis in model building. In the case of very distant homology, these issues are overshadowed by the necessity to correctly map the target sequence onto the conserved regions of the template(s) in the first place. Accordingly, in such case, the correctness of the alignment and optimal use of structural information from the available templates are the most significant determinants of the quality of the final model.

For the fourth round of Critical Assessment of Techniques for Protein Structure Prediction (CASP4) I have submitted models for 13 target proteins. They included both distant and high homology targets, providing a comprehensive test for the modeling approach presented here.

Because of my affiliation with the Prediction Center, where all models were handled, I have felt obliged to ensure the possibility to verify that I have honored the model submission deadlines as did all other prediction groups. To do this, upon depositing a model into the CASP4 database, I would send a copy of the same model along with the date stamp to one of the independent assessors (Manfred Sippl).

In this article, I briefly describe the approach used at CASP4 and present an analysis of the obtained results with emphasis on sequence-structure alignments. Using models for distant homology targets I provide examples of both successful and unsuccessful identification of correct alignment and discuss the underlying reasons for that. Finally, I analyze the effect of using multiple templates

Joint affiliation with Institute of Biotechnology, Graičiūno 8, 2028 Vilnius, Lithuania.

*Correspondence to: Česlovas Venclovas, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94550. E-mail: venclovas@llnl.gov

Received 13 March 2001; Accepted 6 August 2001

and provide results for loop modeling as well as accuracy of side-chain positioning.

METHODS

During the fourth round of "blind" prediction experiment, I have adopted comparative modeling strategy similar to the one successfully used 2 years earlier[1] in CASP3. The main feature of this strategy was to concentrate on the improvement of initial steps in model building, such as sequence-structure alignment and use of the available structural information from related proteins (templates). Although at CASP4 generating and assessing sequence-structure alignments remained the main emphasis in modeling distant homology targets, the actual procedure was significantly modified. A brief description of the alignment step and other major procedures used to generate CASP4 models is provided below.

Deriving and Assessing Sequence-Structure Alignments for Low to Moderate Homology Targets

Initially, using PSI-BLAST[2] with standard parameters, sequence of each target protein was compared with all other sequences in the nonredundant protein sequence database (nr) taken from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). Results of the initial PSI-BLAST search were used for the intermediate sequence search procedure (PSI-BLAS-ISS), a first step toward generating sequence-structure alignment. In this procedure, a set of sequences that bridged the sequence space between target and templates in the initial search results and that were $<60\%$ identical to one another was identified. Each sequence from this set was used in turn as a probe to generate the corresponding PSI-BLAST profile by searching the nonredundant sequence database with a more stringent expectation value cutoff (typically 10^{-10}–10^{-20}). Using the SEALS package[3] and in-house Perl scripts, target-template sequence alignments were extracted from the resulting individual PSI-BLAST profiles and compared. The convergence of the target-template alignments in a particular region was then used as an indicator of the alignment reliability in this region. Thus, if most of the target-template alignments extracted from different PSI-BLAST profiles were identical for a specific region, this region was considered reliably aligned. If several major alternative alignments were present, none of them were considered reliable at this stage. Subsequently, all these alignment variants were tested by building and evaluating the corresponding models. Some of the regions were not aligned within any of the resulting PSI-BLAST profiles, yet they were expected to be structurally conserved on the basis of structure comparisons between the proteins possessing the same fold. In such cases, alignments were derived manually guided by the PSIPRED[4] secondary structure predictions.

As mentioned above, the final selection of the region-specific sequence-structure alignments was done by building and evaluating the corresponding 3D models. In an attempt to make better judgment regarding correctness of the alignment in the questionable regions, in most cases models were built not only for the target protein but also for its close homologues. Evaluation of the 3D models was performed by visual inspection with emphasis on significant structural flaws, such as buried uncompensated charges or hydrogen donors/acceptors and severe steric clashes as detected with the structure verification module of WHATIF.[5] In addition, for models of moderately distant targets and their homologues, a consensus of ProsaII[6] Z-scores was also used.

Generating 3D Structures

Models for alignment evaluation purposes were built either with InsightII (MSI Inc., San Diego, CA) Homology module or with MODELLER4.[7] The final models were always generated with MODELLER to enable automatic combination of multiple templates. Templates for distant homology targets typically were selected from the set of PDB structures identified after a third iteration of the PSI-BLAST search with the target sequence against the nonredundant sequence database. In cases of high homology, templates were chosen from the results of target sequence in comparison with all PDB structures using the Smith-Waterman algorithm[8] implemented in the SSEARCH.[9] If a large number of templates were available, only several of them (typically 2–6) representing structural variation within the respective protein family or superfamily were selected.

For the distant homology targets, all of the modeled loops were also generated automatically with MODELLER. For moderate and high homology targets, the coordinates for some of the loop regions were assigned from suitable fragments found in PDB structures. The preference was given to the proteins evolutionary related to the target. In the absence of suitable fragments from homologous structures, loop regions were assigned dominant fragment conformations. Side chains for the obtained models were positioned by using a backbone-dependent rotamer library implemented in SCWRL.[10] If after this step there were remaining severe side-chain clashes, they were abated by manual rotamer positioning. No other model refinement procedures were applied.

RESULTS AND DISCUSSION

Experimental structures for 12 of 13 attempted target proteins have been solved in time to assess the quality of models at the CASP4 Asilomar meeting.[11] Sequence identity between these proteins and the closest known 3D structures ranged from 58% down to 13%, representing the entire range of prediction difficulty among the targets considered in comparative modeling assessment.[12]

Sequence-Structure Alignments

Modeling results with the emphasis on sequence-structure alignments are summarized in Figure 1. For reference, the figure also includes the average values for models submitted by all predictor groups. Data regarding model structures in Figure 1 were extracted from the publicly available CASP4 numerical evaluation database at the Prediction Center CASP4 web site (http://PredictionCenter.

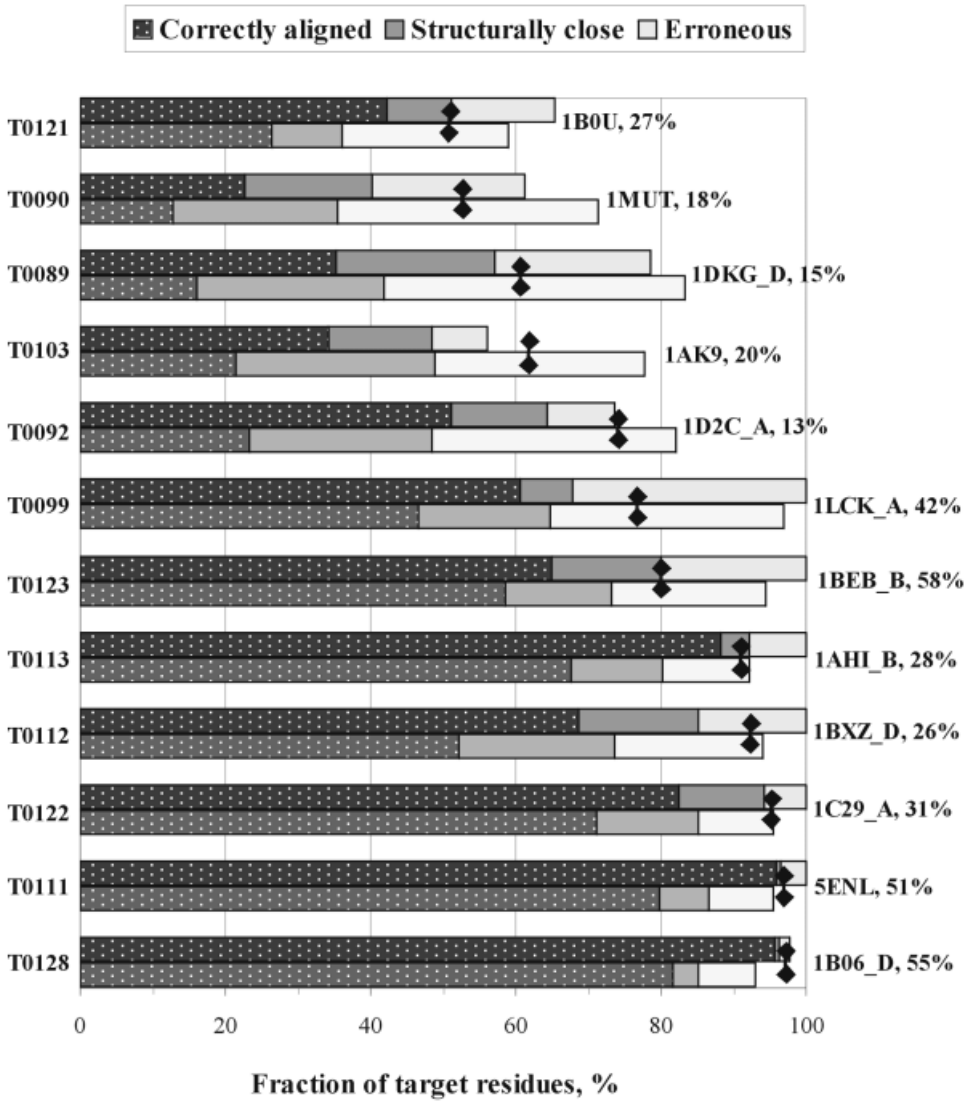

Fig. 1. Summary of model quality. The upper and lower stacked bars represent, respectively, values for my models and the average values calculated for highest confidence models ("Model 1"s) from all groups. "Correctly aligned" is a fraction of residues in the model that were correctly aligned; "structurally close" is an additional fraction of structurally equivalent residues for which alignment was wrong, and "erroneous" are the remaining residues in the model. Diamonds indicate the fraction of structurally equivalent residues identified upon superposition of target and structurally closest template. This value is directly comparable with the sum of "correctly aligned" and "structurally close" fractions calculated for the submitted models. PDB codes of the templates and sequence similarity within superimposed regions are also indicated. The data are arranged from top to bottom according to the increase in fraction of structurally equivalent residues between target and template. Low fraction of equivalent residues in the case of T0121 is due simply to the absence of C-terminal domain in the template structure.

llnl.gov/casp4) as derived from sequence-independent target-model superposition generated with the LGA program.[13] In this figure, I also attempt to estimate the results that potentially could be achieved by using a "perfect" (structure-based) target sequence alignment with the structurally closest template. The estimates are based on the target-template structural comparison using LGA, for the sake of consistency applying the same 4 Å equivalence cutoff that was used to obtain data for models.

For the CASP4 experiment, I put forward the same goals as for the comparative modeling applications in "real-life" projects, which is to maximize the fraction of accurately modeled structure without proliferating errors at the same time. In practice, this meant concentrating on the correctness of the model within the structurally conserved regions and omitting regions that were expected to differ significantly and would most likely substantially increase the amount of errors. It should be emphasized that both automatic and human CASP4 assessments were more forgiving in this respect, because they focused only on the positive aspects of predictions and disregarded the extent of errors. It is not surprising that because of the stringent

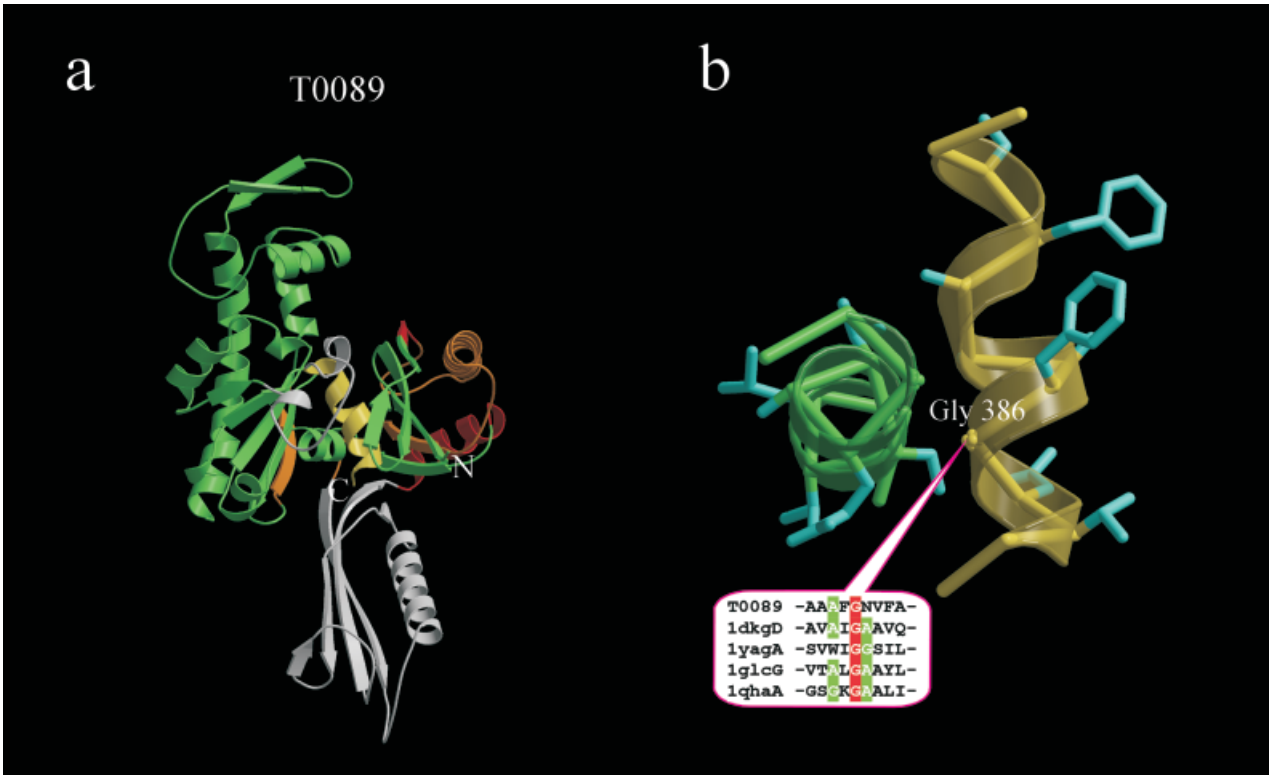

Fig. 2. Sequence-structure alignment for target T0089. **a:** The X-ray structure of the target (PDB code: 1E4F). Color coding is as follows: green, correctly aligned regions; orange, misaligned region; red, alignment errors that could be avoided if alignment from a standard PSI-BLAST search were used; yellow, C-terminal helix, aligned correctly due to 3D assessment; gray, insertions, not present in parental structures. **b:** Close-up view of C-terminal helix interaction with pseudosymmetry related helix. Position, corresponding to Gly 386 in T0089, is absolutely conserved in related structures. Cartoon representation of structures in this and other figures was generated with Molscript[19] and Raster3D.[20]

nature of my modeling objectives, the resulting models for the difficult targets (T0089, T0090, T0092, and T0103) are less complete than the average. Despite that, like models for the rest of targets, they all have larger than the average fraction of correctly aligned residues. For two of the targets (T0089 and T0092), this fraction exceeds the average more than twice.

Although modeling results are consistently better than the average, there is still a lot of room for improvement. Although regions, structurally equivalent between target and template, in most models are reproduced successfully, correct alignment for distant homology targets is still significantly below the estimated achievable level (Fig. 1). Because the alignment approach reported here is a mosaic of sequence and structure methods, the question is what was the contribution of each step in generating correct alignments, what were the causes of remaining alignment errors, and what could be the possible solution to avoid them?

The first step in obtaining sequence-structure alignments, PSI-BLAST-ISS, served a twofold purpose: (a) to estimate region-specific reliability of the alignment and (b) to provide candidate alignments for testing as 3D models in questionable regions. Although more extensive testing is required, the estimation of the region-specific alignment reliability seems to have worked well for CASP4 targets. All alignment errors observed in the final models were located only in the unreliable regions that within the PSI-BLAST-ISS results either displayed significant diversity of alignment variants or were not aligned at all. In such unreliable regions, alignment errors were caused by either absence of the correct alignment within the set of tested variants or failure to select the correct alignment from a number of alternatives by evaluating corresponding models. Below, using data from models for three distant homology targets (T0089, T0092, and T0103), I illustrate such alignment problems and analyze the reasons of why they were or were not solved. Because PSI-BLAST was an important component of the alignment approach, I compare resulting alignments with those that would have been obtained if alignments from standard PSI-BLAST search at the time of CASP4 experiment were used to generate models.

T0089 (cell division protein FtsA, T. maritima)

The structure of this target protein[14] resembles that of the actin family, including actin, heat-shock cognate (Hsc) protein, hexokinase, and glycerol kinase. This was the only target where two of the regions that got misaligned in a model submitted to CASP4 [colored in red in Fig. 2(a)] would have been correctly aligned by PSI-BLAST alone (with default parameters). Although the correct alignment appeared among variants produced with the PSI-BLAST-ISS procedure, it was dismissed by the structure-based

selection. Another misaligned region coincided with the helix-loop-strand motif connecting points of deletion of one subdomain (≈40–70 a. a.) and insertion of another subdomain (≈80 a. a.) with respect to either actin or heat-shock protein. In this case, none of the variants suggested by PSI-BLAST-ISS was even close to the correct alignment. It is of interest that, according to the numerical evaluation data available from the Prediction Center's database, not a single T0089 model produced by any predictor group had the correct alignment for this region.

At the same time, the C-terminal helix for this target provides an excellent example of how even the tiny structural details may be significant in identifying the correct alignment. For this helix, PSI-BLAST-ISS suggested two major alignment variants, one of which corresponded to the correct alignment. It is of interest that this particular helix is packed against another helix at ≈90° angle [Fig. 2(b)]; each of these helices are part of the pseudosymmetry related RNase H-like domains that are common for the actin family. At the point of contact between the two helices, the backbone of C-terminal helix comes very close to the interacting helix. Structural analysis reveals that in the position of contact even the smallest side chain (Ala) would produce steric clashes, suggesting that the presence of glycine in this position is not accidental. Indeed, the corresponding position is occupied by glycine throughout the entire actin family, confirming its significance. That allowed selection of the correct alignment with high confidence, because in the alternative variant, the corresponding position was aligned with residue other than glycine.

T0092 (protein HI0319, H. influenzae)

The sequence of this "hypothetical" protein displays distant but significant similarity to S-adenosyl-L-methionine-dependent methyltransferases. Most of them have structurally conserved cofactor binding domain, formed by a seven-stranded β-sheet sandwiched by α-helices and an additional variable domain, dependent on methylation substrate. All PSI-BLAST-generated alignments with related methyltransferases of known structure terminated after the fifth strand, suggesting that, in T0092, this is the boundary with the variable domain of a novel structure. At the same time, secondary structure prediction data, combined with the consensus structure of methylase fold, indicated that T0092 C-terminus should form a conserved helix-β-hairpin motif characteristic to most methyltransferases. Consensus results of the assessment of 3D models for this target and its several homologuess, based on alignment variants suggested by PSI-BLAST-ISS, enabled to correctly map the sequence onto all conserved secondary structure elements preceding insertion domain (Fig. 3). By comparison, PSI-BLAST alone would have misaligned a helix and a strand in the same region. Sequence-structure mapping for the C-terminal motif was less successful, with sequence of the β-hairpin being shifted by one residue. It appears that, although this motif is common to many methyltransferases, it is structurally less conserved and even completely missing in some structures. In addition, most of the side-chain interactions within this helix-β-hairpin motif are local. That made it almost impossible to derive any additional structural constraints, originating from the other regions of the domain, that could be used to discriminate correct and erroneous alignments.

T0103 (pepstatin-insensitive carboxyl proteinase, Pseudomonas sp.)

This protein is related to the subtilisin clan of serine proteases[15] that all have a single domain composed of seven β-strands flanked by a number of helices. In general, both defining the structurally conserved regions and producing the sequence-structure alignment was not trivial for this target. However, here I focus only on the alignment of a single helix, which, as judged by similarity to related serine proteases, is involved in active site formation (Fig. 4). Despite every attempt to identify correct sequence-structure mapping by exploring a large number of candidate alignments, the sequence of this helix in the resulting model has been shifted by one helical turn in respect to the experimental structure. Why was the structurally correct alignment missed in this case? Retrospective analysis shows that the correct alignment for this helix was not among variants produced by PSI-BLAST-ISS. However, because there was no dominant variant, a number of alignments, including the correct one, were systematically assessed by 3D model evaluation. Yet none of the alignments tested for this helix were acceptable within the framework of protein structures from subtilisin superfamily. Figure 4 illustrates why. Although structural arrangement of this helix is conserved, both surrounding structural environment and the amino acid sequence of the helix itself undergo significant changes. Compared with the corresponding helix in subtilisin, the place of His64 from the catalytic Asp-His-Ser triad is occupied by Glu80, which is no longer coordinated by Asp from the adjacent strand (Asp32). Instead, Glu80 is coordinated by Asp, which is located within the same helix (Asp84) in place of a hydrophobic residue present in subtilisin. However, the most dramatic change corresponds to the substitution of Gly (subtilisin) with Trp81 (T0103). To accommodate a bulky Trp side chain, the region of the protein chain, adjacent in space to this point substitution, moves away up to ≈5 Å. Immediately following the shifted region, a regular helix is formed in T0103 instead of irregular conformation for the equivalent region in subtilisin, significantly modifying the environment. Therefore, it appears that, without anticipation of the rearrangements within the structural environment of the active site helix, it was impossible to produce a structurally sound model even having the correct sequence-structure alignment.

The above examples show that structure-based assessment of candidate alignments proved to be effective in some difficult cases (e.g., C-terminal helix of T0089), whereas in some (e.g., C-terminal motif of T0092 and active site helix in T0103) it was not. It seems that structural evaluation of the alignment variants as 3D models is most useful when the structure for the target is highly conserved even if sequence similarity is very low. The experience with modeling CASP4 targets also sug-

gests that the sensitivity of structure-based evaluation of alignments can be increased by generating and evaluating models not only for the target sequence but also for its close relatives. This enables detection of some alignment errors that may not show up as significant flaws in the 3D models for some of the family sequences but may do so for others. Certainly, as illustrated by the alignment of T0103 active site helix, this is not always sufficient. In such cases, the ability to move away from the rigid template might be essential for identification of the correct alignment.

Fig. 3. Sequence-structure alignment for the target T0092. Color coding is the same as in Figure 2. The C-terminal helix-β-hairpin motif following long insertion is indicated with a broken line.

Among issues related to the structure-based evaluation of alignments is how to ensure adequate sampling of variants to be tested, because exhaustive assessment of all possible variants is hardly feasible. PSI-BLAST-ISS was introduced as an attempt to provide a limited but representative set of alignment variants, which at the same time addresses the issue of reliability for each of the considered regions. However, because PSI-BLAST-ISS is nothing else but a way of using PSI-BLAST results, it suffers from the same flaws as PSI-BLAST. Specifically, long insertions or deletions (e.g., T0089) cannot be handled properly because of large gap penalties, resulting in a set of candidate alignments that may not even be close to the correct variant. Perhaps including information regarding structural variability within proteins of corresponding fold might partially address this issue.

Template Selection and Use

Presently, a structural similarity between target and template(s) is the major determinant of the upper margin of model accuracy. Although in general structural similarity correlates with sequence similarity, this relationship becomes increasingly fuzzy as sequence homology decreases (e.g., see Fig. 1 in Venclovas et al.[16]), making optimal choice of the template nontrivial. To eliminate this problem, I used multiple templates whenever they were available. That was one of the reasons to choose MODELLER, a program that is capable of automatically taking into account structural data from multiple templates. After the experimentally determined target structures became available, I decided to examine whether a combination of multiple templates always led to the improvement over use of individual template structures. To answer this

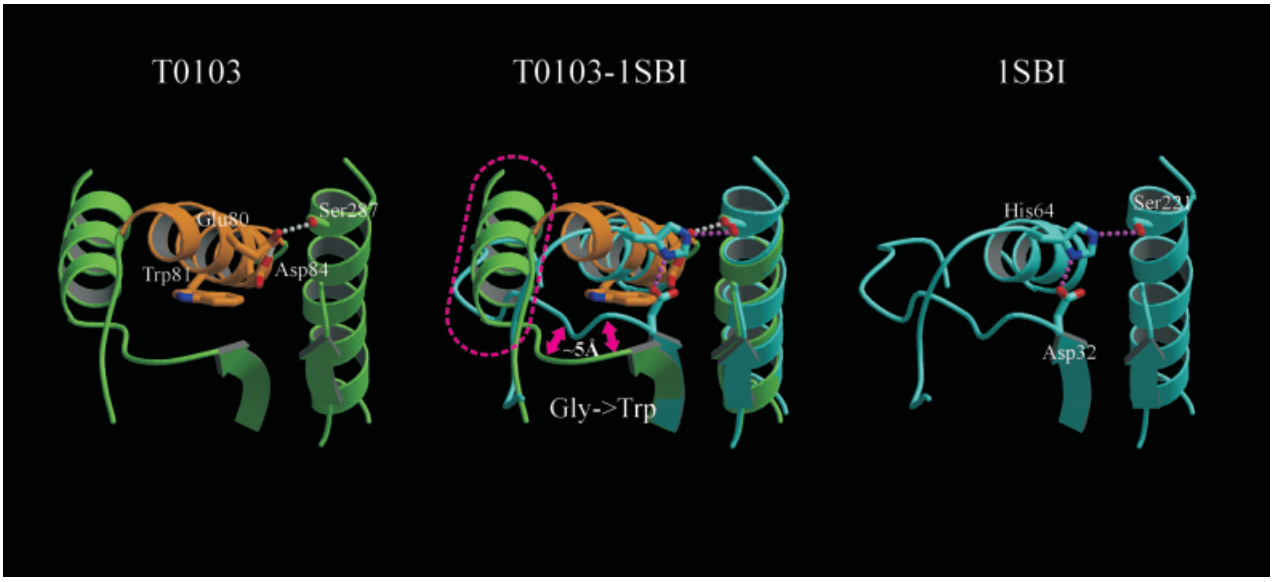

Fig. 4. Structural changes in the vicinity of the misaligned active site helix (orange) in T0103 (1GA6) compared to subtilisin (1SBI). Magenta arrows indicate shifted region in T0103 associated with the Gly→Trp substitution. A broken line indicates regular helix formation in T0103 in place of the corresponding region of irregular conformation in subtilisin. Hydrogen bonding between active side residues in subtilisin and between putative active site residues in T0103 is shown, respectively, as magenta and gray dots.

question, I generated models using the same alignments as for the models submitted to CASP4 but based on individual structures from the template set. The relative quality of models was then compared by using target-model RMSD values for all Cα atoms. It turned out that only in one case (T0123), the CASP4 model built by using two templates was not better than models based on any of two individual templates. In case of four targets, CASP4 models based on the combined set of templates were better than models based on most individual templates but were still slightly worse than the models produced by using the best template from the set. For the remaining four targets, models based on multiple template structures were closer to the target than the models based on any individual template. The largest gain (0.4 Å RMSD) was observed for T0092 and T0113, and the improvement is considerable compared with the total RMSD values: 2.9 Å/167Cα atoms for T0092 and 2.4 Å/255 Cα atoms for T0113.

These results suggest that use of multiple templates has a potential to significantly improve models compared with the ones based on a single template. However, this potential is not always realized, and consistency for optimal combination of multiple structures has yet to be achieved.

Loop Building, Orienting Side Chains

In case of distant homology targets, very long insertions were skipped from modeling, whereas the remaining loops were modeled automatically with MODELER. Only in models for five high homology targets, some loops were assigned an explicit conformation from the suitable fragments found in the PDB structures. For the models of four targets, the explicit loop modeling did not produce any significant change in overall quality of models (RMSD for all Cα atoms in the model changed <0.03 Å) compared with an automatic loop assignment. Only for target T0113, manual assignment of four loops did result in a detectable improvement of the model (overall RMSD decreased by 0.13 Å). However, even this largest improvement appears to be small compared with the effects observed due to the selection and/or combination of structural templates and those due to alignment errors.

Accuracy of side-chain rotamers, expressed as percentage of correct (within $\pm 30°$) χ_1 angles, ranged from 48% for T0090 to 68% for T0128. If only buried side chains were considered, the accuracy of all models increased by $\approx$10% (ranging from 57 to 78%). Side-chain prediction accuracy for target T0103, for which the step of rotamer repositioning with SCWRL was skipped, was significantly worse, with only 38% of χ_1 correct and did not improve if only buried side chains were considered. Most surprisingly, models for two targets, one displaying the most distant sequence homology (T0092, 13% sequence identity) and the other displaying the highest sequence similarity (T0123, 58% sequence identity) among comparative modeling targets, were very close in side chain prediction accuracy. T0092 had 53% of χ_1 angles correct, whereas T0123 had just 54%. Although this data may seem surprising, it is in agreement with the observation that structural, rather than sequence, conservation is a determining factor for side-chain prediction accuracy (e.g., Refs. 17 and 18). Despite a huge difference in sequence homology, both targets display approximately the same level of structural similarity to their respective templates (Fig. 1).

CONCLUSION

Sequence-structure alignments are still the dominant source of errors in modeling distant homology targets. There seem to be two major causes for alignment errors in the structurally conserved regions: (a) unanticipated large insertions or deletions flanking these regions and (b) variations in the surrounding structural environment. Significant improvements in addressing these issues are yet to be seen.

Combination of structural information from multiple templates can significantly increase model accuracy. Because availability of multiple templates due to constant increase of structural databases is going to be a rule rather than an exception, the development of methods that can optimally combine multiple template structures might be very fruitful. Progress in this area would also lessen secondary problems of model construction such as loop modeling and side-chain positioning.

ACKNOWLEDGMENTS

I thank the structural biologists who provided target proteins for CASP4 experiment and K. Fidelis for critically reading the manuscript. This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

REFERENCES

1. 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.
2. 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.
3. Walker DR, Koonin EV. SEALS: a system for easy analysis of lots of sequences. Proc Int Conf Intell Syst Mol Biol 1997;5:333–339.
4. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.
5. Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:29, 52–56.
6. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins 1993;17:355–362.
7. Šali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
8. Smith TF, Waterman MS. Identification of common molecular subsequences. J Mol Biol 1981;147:195–197.
9. Pearson WR. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. Genomics 1991;11:635–650.
10. Bower MJ, Cohen FE, Dunbrack RL Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. J Mol Biol 1997;267:1268–1282.
11. Moult J, Hubbard T, Zemla A, Fidelis K. Critical assessment of methods of protein structure prediction (CASP): round IV. Proteins 2001; Suppl 5:2–7.
12. Tramontano A, Morea V, Leplae R. Analysis and assessment of comparative modeling predictions in CASP4. Proteins 2001; Suppl 5:22–38.
13. Zemla A. LGA program—a method for finding 3-D similarities in

protein structures. Accessed at http://PredictionCenter.llnl.gov/local/lga/lga.html. 2000.
14. van Den Ent F, Löwe J. Crystal structure of the cell division protein FtsA from Thermotoga maritima. EMBO J 2000;19:5300–5307.
15. Wlodawer A, Li M, Dauter Z, Gustchina A, Uchida K, Oyama H, Dunn BM, Oda K. Carboxyl proteinase from Pseudomonas defines a novel family of subtilisin-like enzymes. Nat Struct Biol 2001;8:442–446.
16. Venclovas Č, Zemla A, Fidelis K, Moult J. Some measures of comparative performance in the three CASPs. Proteins 1999;Suppl: 231–237.
17. Chung SY, Subbiah S. How similar must a template protein be for homology modeling by side-chain packing methods? Pac Symp Biocomput 1996:126–141.
18. Martin AC, MacArthur MW, Thornton JM. Assessment of comparative modeling in CASP2. Proteins 1997;Suppl:14–28.
19. Kraulis PJ. Molscript—a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24: 946–950.
20. Merritt EA, Bacon DJ. Raster3D: photorealistic molecular graphics. Methods Enzymol 1997;277:505–524.