Comparative Modeling in CASP6 Using Consensus Approach to Template Selection, Sequence-Structure Alignment, and Structure Assessment

Česlovas Venclovas* **and Mindaugas Margelevičius**
Institute of Biotechnology, Vilnius, Lithuania

ABSTRACT **Along with over 150 other groups we have tested our template-based protein structure prediction approach by submitting models for 30 target proteins to the sixth round of the Critical Assessment of Protein Structure Prediction Methods (CASP6, http://predictioncenter.org). Most of our modeled proteins fall into the comparative or homology modeling (CM) category, and some are fold recognition (FR) targets. The key feature of our structure prediction strategy in CASP6 was an attempt to optimally select structural templates and to make accurate sequence–structure alignments. Template selection was based mainly on consensus results of multiple sequence searches. Likewise, the consensus of multiple alignment variants (or lack of it) was used to initially delineate reliable and unreliable alignment regions. Structure evaluation approaches were then used to identify the correct sequence–structure mapping. Our results suggest that in many cases use of multiple templates is advantageous. Selecting correct alignments even within the context of a three-dimensional structure remains a challenge. Together with more effective energy evaluation methods the simultaneous relaxation/refinement of a "frozen" backbone inherited from the template is likely needed to see a clear progress in tackling this problem. Our analysis also suggests that human input has little to contribute to automatic methods in modeling high homology targets. On the other hand, human expertise can be very valuable in modeling distantly related proteins and critical in cases of unexpected evolutionary changes in protein structure. Proteins 2005;Suppl 7:99–105.** © 2005 Wiley-Liss, Inc.

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

INTRODUCTION

Currently the growth of numbers of experimentally determined protein structures is heavily outpaced by the growth of protein sequence databases, and this trend is not likely to change in foreseeable future. Inescapably, an important role in reducing disparity between the volume of sequence and structural information belongs to computational methods. Of these, comparative modeling often is the method of choice when it comes to structural characterization of protein sequences that could be related (even distantly) to known structures. Depending on accuracy, protein models can be used to address a range of biologically relevant questions by both experimental and computational approaches. In comparative modeling the accuracy tends to decrease as the evolutionary distance between the sequence being modeled (target) and the structural template increases. One way to address that is to identify unreliable regions of the model and ignore them. However, all the more desirable goal of any protein prediction method including comparative modeling both in CASP setting and in real-life projects is to model accurately and reliably as many regions of the structure as possible.

According to the experience of one of us (Č.V.), gained through a fairly successful participation in three previous CASP experiments,[1–3] there are two factors that affect the accuracy of the template-derived model most. Sequence–structure mapping (alignment) errors is the single most detrimental problem in distant comparative modeling. In medium and high homology cases, when alignment is no longer a problem, an optimal selection of structural templates becomes the decisive factor of the model accuracy. Therefore, during CASP6 we decided to concentrate almost exclusively on these two problems and find out whether our updated approach is successful in addressing them. Another obvious goal for participating in CASP6 was to compare the performance of our approach relative to others. Because of the large increase in numbers of fully automated methods in recent years, we also were interested in finding out the range of prediction targets and modeling problems, for which the value of human intervention is currently outweighing the benefits of automation.

METHODS

Selection of Structural Templates

PDB templates were identified by running either BLAST or PSI-BLAST[4] searches against the PDB sequence data-

Grant sponsor: Howard Hughes Medical Institute; Grant sponsor: the 6th European Community Framework Programme.

*Correspondence to: Česlovas Venclovas, Institute of Biotechnology, Graičiūno 8, LT-02241 Vilnius, Lithuania. E-mail: venclovas@ibt.lt

Received 15 April 2005; Accepted 31 May 2005

Published online 00 Month 2001 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20725

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint verison.

base or nonredundant NCBI sequence database, respectively. If no significant matches to PDB entries were detected then consensus results reported by 3D-Jury[5] at http://bioinfo.pl/ and the GeneSilico fold recognition metaserver (http://genesilico.pl/meta)[6] were consulted. In those cases we only modeled targets, for which fold assignment was made with high certainty. We used multiple templates whenever they were available unless one of the templates was expected to be significantly closer to the target structure than the others. Selection of the structural templates was usually based on the consensus results obtained with sequence searches. Templates detected most frequently during multiple PSI-BLAST searches initiated with target homologs were considered to be better representatives of the target structure and therefore stronger candidates to be selected.

Sequence–Structure Alignments

For high homology targets, where structural template(s) were among closely related sequences, alignments were derived directly from BLAST or PSI-BLAST results. Manual adjustments guided by template structures were sometimes introduced only to better position insertions/deletions. For distant homology targets, two methods were used to generate and make preliminary assessment of the alignment confidence in a region-specific manner. In the first method, results of an initial PSI-BLAST search were used in our intermediate sequence search procedure (PSI-BLAST-ISS).[2] In this procedure, a set of sequences that bridge sequence space between the target sequence and template(s) are used to initiate additional PSI-BLAST searches against the nonredundant sequence database. Target–template sequence alignments are then extracted from search results and their consistency is analyzed. For regions where one dominant alignment variant is produced, the alignment is considered reliable, while the regions where the consistency of target–template alignment is lacking are deemed unreliable. In the second method, publicly available 3D models for a particular target that were submitted to CASP6 by automatic servers were each superimposed with one of the templates using DaliLite.[7] Next, the structure-based multiple sequence alignment between the template and model sequences was constructed from obtained pairwise superpositions. The region-specific alignment reliability was then assessed as in the first method. Results by both methods were contrasted and consensus regions were considered to be reliably aligned. For the remaining regions alternative alignment variants were evaluated at the level of 3D models. Models based on these alternative alignments were assessed by several methods including ProsaII[8] profiles and *Z*-scores, Verify3D profiles,[9] and visual inspection. Verify3D was a new addition in CASP6 in hopes to make assessment of models even more rigorous than in earlier CASPs. We have used ProsaII *Z*-scores as a main numerical criterion for ranking models based on alternative alignments and alternative placement of insertions/deletions. We also used ProsaII *Z*-scores to estimate the quality of our models relative to CASP6 server models. The targeted goal was to produce a model that according to the ProsaII *Z*-score would fare better than any server-generated model.

Construction of Models

Three-dimensional models both for initial evaluation purposes and for submission were generated from given sequence–structure alignments automatically with MODELLER.[10] In most cases side chains were rebuilt with SCWRL.[11] Manual intervention was kept to a minimum and no energy minimization procedures were used.

RESULTS AND DISCUSSION

During the CASP6 experiment we have submitted models for 30 prediction targets. All of the targets were modeled using structural templates. Although we have made some predictions for targets classified by the assessors as fold recognition (FR) or even new fold (NF), most of our modeled targets fall into the comparative modeling (CM) category. The summary of all our assessed predictions are provided in Table I. The table gives the structural template-centered view of prediction quality. Comparison of values in fifth and seventh columns indicates how successful we were in both utilizing the template structures and correctly aligning the target sequence with templates. If values for sequence-dependent (SD) superposition of target and model (seventh column) are better than the sequence-independent (SI) superposition of the target and the template (fifth column), it means that the model is structurally closer to the target than the template even if the one-to-one sequence correspondence is taken into account. In many cases the values are comparable. There are some models that produce superposition values better than those for the template, but also there are models that are clearly worse. Most of the latter modeling targets are distantly related to the templates (FR or CM/hard categories), and their poor quality can be attributed to failures in producing correct alignment and/or identifying structurally conserved parts of templates.

Multiple Templates Versus Single Template

In most cases we have used more than one structural template to construct a model. An obvious question is whether the models derived from multiple templates are structurally closer to the target than those which would have been derived from a single best template. For this we analyzed comparative modeling targets, which were modeled using multiple templates. We used sequence–structure alignments, from which submitted CASP6 models were derived, to generate models based on each template individually. Then all the models including the submitted one were compared with the target structure using the sequence-independent mode (at 5-Å distance cutoff) of the LGA program.[12] The number of structurally equivalent residues and RMSD values of Cα atoms were used to judge which of the models is structurally closest to the target. It turned out that the model based on multiple templates was not always the best according to this measure. In some cases it would have been better to use just a single closest

TABLE I. Summary of CASP6 Predictions

Target/Domain	Category	Templates used	TN	SI T-P eqv/RMS	T-P seq_id, %	SD T-M, eqv/RMS
T0196	CM/easy	**1jny_A,** lexm_A	89	79/1.5	33	81/1.8
T0199_1	CM/hard	**1in4_A,** 1jhf_A, 1j5y_A, 1mkm_A	74	67/2.0	21	71/1.9
T0200	CM/hard	1ush	255	202/2.5	17	191/3.0
T0201	NF	**1fno_A,** 1fxl_A	94	75/3.2	8	53/3.5
T0204	CM/easy	**1guq_A,** 1gup_A	297	276/1.9	26	281/2.0
T0208	CM/hard	**1dxi_A,** 1k77_A, 1i60_A	344	234/2.6	12	230/2.6
T0211	CM/hard	**1eut,** 1jhj_A, 1czs_A, 1kex_A, 1k12_A	136	126/1.8	22	126/2.0
T0213	FR/H	1t62_A	103	80/2.8	14	75/2.6
T0222_1	CM/hard	**1rzm_A,** 1jcx_A, 1d9e_A	264	230/2.0	14	230/2.2
T0223_1	CM/hard	1bkj_A	114	110/2.0	15	84/1.8
T0223_2	FR/H		92	76/1.7	18	32/1.1
T0228_1	FR/H	1qpo_A	157	98/2.7	15	64/2.6
T0228_2	FR/H		235	142/2.8	11	98/3.1
T0229_1	CM/easy	**1ml8_A,** 1ukk_A, 1nye_A	24	24/0.8	29	24/0.9
T0229_2	CM/easy		102	95/1.9	37	98/2.2
T0231	CM/easy	1v6f_A	137	135/1.3	80	135/1.2
T0232_1	CM/hard	**1f3a_A,** 1gul_A	81	74/1.7	24	78/2.2
T0232_2	CM/hard		146	118/2.3	12	116/2.5
T0233_1	CM/easy	**1v8g_A,** 1kgz_A, 1khd_D	66	62/1.5	23	66/1.3
T0233_2	CM/easy	1o17_A	265	253/1.4	45	249/1.5
T0234	CM/hard	**1dnl_A,** 1nrg_A, 1flm_A	135	114/2.2	16	106/2.1
T0235_1	CM/easy	**1nbf_A,** 1nb8_A	309	271/1.8	28	270/2.1
T0235_2	FR/A		43	n/a	n/a	22/2.1
T0244	CM/easy	**1iin_A,** 1h5r_A, 1fxo_A	296	232/2.0	23	225/2.3
T0247_1	CM/easy		150	141/1.6	26	129/1.7
T0247_2	CM/easy	1pj5_A	135	125/1.5	26	133/1.7
T0247_3	CM/easy		76	71/1.7	21	71/1.8
T0266	CM/easy	**1dbx_A,** 1vki_A	150	147/1.7	24	150/2.0
T0267	CM/hard	**1tiq_A,** 1i12_A	174	160/1.9	18	161/2.0
T0274	CM/easy	**1i0r_A,** 1rz0_A, 1eje_A	156	145/1.5	23	143/1.6
T0275	CM/easy	**1mjh_A,** 1jmv_A, 1tq8_A	135	126/2.1	29	125/2.0
T0279_1	CM/hard	1jr2_A	127	114/2.3	20	121/2.5
T0279_2	CM/hard		121	115/2.4	10	115/2.6
T0282	CM/easy	**1gq6_A,** 2cev_A, 1pq3_A	232	264/2.0	23	261/2.0

Target/Domain: prediction target or domain.
Category: assessor-based classification of targets. Comparative modeling (CM/easy and CM/hard), Fold recognition–homologous structure (FR/H), analogous structure (FR/A) and New fold (NF).
Templates used: PDB structures (chains) that were used to build models for a given target; templates indicated in bold are structurally closest to the target among those listed.
TN: number of residues in a target structure
SI T-P eqv/RMS: the number of structurally equivalent residues between target and parent (template) and the corresponding RMSD value. Values were derived from sequence-independent (SI) superposition with LGA using 5 Å distance cutoff.
T-P seq_id, %: sequence identity between target and parent (template) derived from the structure superposition
SD T-M, eqv/RMS: the number of structurally equivalent residues between target and model and the corresponding RMSD value. Values were derived from sequence-dependent (SD) superposition with LGA (5 Å distance cutoff).

template. The impact of multiple templates on model quality relative to other predictor groups could be seen most clearly for medium/high homology (CM/easy) targets. For example, models for T0235, T0274, T0275, and T0282, where a combination of templates led to improvement over any individual template used, all appear among top five models according to the GDT_TS score. In contrast, our model for T0229, for which a single template, would have been a better choice, although it has a correct alignment fare much worse in comparison to many other groups. Thus, our results suggest that it would be too simplistic to claim that using multiple templates rather than a single one is always better. It depends on a particular modeling case. However, given the difficulty of identifying the best template beforehand, the use of multiple templates offers an increased chance that at least one of the selected templates is among the best.

Sequence–Structure Alignments

Sequence–structure alignments were one of our major priorities, and we were able to correctly align many difficult regions where most of predictor groups failed. Good examples are targets T0208 and T0211, which are discussed in detail below. Conversely, the majority of our alignment errors were made in regions where usually less than 10–15% of predictor groups were correct. An encouraging observation is that essentially all of alignment errors that appear in structurally conserved regions of our

models were not unexpected. The exception is T0266, an easy CM target, where an alignment error was inadvertently introduced by a human error. Thus, both PSI-BLAST-ISS results and structure-based alignments of public server 3D models were effective in delineating unreliable alignment regions. However, it is frustrating to see that a successful detection of problematic regions has not always led to a successful identification of the correct sequence–structure mapping. From the experience of earlier CASPs it was clear that in many difficult to align regions sequence pattern matching methods are virtually useless, and that assessment of structures using energy-based approaches might be more promising.[2,3] In CASP6, to better discriminate between correct and incorrect alignments within the context of 3D structure, we added Verify3D profiles to our usual evaluation procedure that includes ProsaII and visual assessment. However, having several methods of evaluating alternative alignments sometimes did complicate things, because it was not always possible to reach the consensus. In such cases the selection of alignment variants had to rely on expert human judgment. In the context of overall CASP6 alignment results we think that more effective local structure evaluation methods based on physicochemical properties perhaps complemented with simultaneous relaxation/refinement of inherited "frozen" backbone are needed.

Manual Modeling Versus Automatic Servers

Because during CASP6 we often used models produced by servers as our self-assessment baseline, it was interesting to compare our results with results from automatic methods. Compared to the best model obtained by any of the participating servers our models trail slightly behind in the "CM/easy" category (the average difference in GDT_TS[13] is −1.0%). However, in the "CM/hard" category, on average, our models outscore the best server-generated models by 1.2% in GDT_TS. If we consider best models from all human groups, then humans outperform servers in both categories, but the trend remains similar. The gap between humans and servers is two times narrower for "CM/easy" targets (T0240 excluded because of classification uncertainty) compared to "CM/hard" (the average difference of 2.1% and 4.5% in GDT_TS values, respectively). This observation suggests that servers are clearly catching up in performance for proteins that have close relatives with known structure, but human input still makes a difference when it comes to modeling distantly related proteins.

Examples of a Beneficial Human Input (T0208, T0211)

Target T0208 is a putative mannonate dehydratase (a member of UxuA protein family). Its structure has been solved by Northeast Structural Genomics Consortium (PDB code: 1tz9). T0208 has a TIM-barrel fold, one of the most abundantly represented folds. Yet, even some of the closest structural templates, such as 1i60 or 1k77, share just above 10% of identical residues with T0208. An unusual structural feature of this prediction target is a long insertion (~65 residues) following the fourth β-strand of the barrel and protruding from the barrel scaffold (Fig. 1). None of the available modeling templates had such an insertion in the corresponding location. As soon as we started working on modeling this target we noticed that PSI-BLAST cannot produce alignments for the N-terminal region. Yet the relationship with TIM-barrel proteins strongly suggested that the missing part of the $(\beta/\alpha)_8$-barrel scaffold must be present within the N-terminal half of the sequence. The inability to align the N-terminal region immediately raised our suspicions about the presence of a long insertion interfering with sequence alignment algorithms. An obvious deviation from a nicely repeating β–α pattern in the predicted secondary structure coincided with the presumable insertion strongly supporting our suspicions. Once we removed the insertion, an alignment could be easily produced with structural templates for the whole length of computationally "engineered" target sequence. Given the low-sequence similarity and structural divergence the final model can be considered fairly accurate with no *bona fide* alignment errors. Although the last α-helix looks as if it is misaligned, in fact, it is just shifted along the helical axis by half of a turn. This shift appears to be due to similar shifts in two out of three templates used. Thus, for T0208 the unsophisticated human intervention step helped us to solve a serious alignment problem, which turned out to be a major culprit in models submitted by all but a handful of predictor groups. We also attempted to model the long insertion in T0208, but were much less successful. The insertion was modeled using guidance from secondary structure prediction and by analogy to the structure of the C-terminal overhang making extensive interchain contacts in a dimer of one of the templates (1dxi). However, the T0208 crystal structure revealed a differently organized dimer having a different orientation of the insertion.

T0211, the protein having galactose-binding domain-like fold, is another structural genomics target (PDB code: 1tvg). The structure has nine β-strands organized into a jelly-roll sandwich formed by two β-sheets (Fig. 2). The biggest challenge in this case involved generating correct sequence–structure alignment for the β-strands forming the edges of the sandwich (β1, β6, and β7). The alignment data for T0211 at the CASP6 Web site (http://predictioncenter.org/casp6/) indicates that these three β-strands were a common problem. The C-terminal β-strand (β9) was also error-prone, but to a slightly lesser degree. The edge β-strands are, in general, notorious for causing alignment problems,[3] but these three were particularly challenging because of the heterogeneity of small local structural features in numerous related structures that could be used as templates. These features include a β-bulge within the β1-strand, another β-bulge or even a small loop within β6 and an insertion of varying length following β7. None of the five templates that we used had all of these features same as in the target structure. The PSI-BLAST-ISS procedure immediately suggested that the sequence–structure alignment in these three regions is not reliable. Therefore, instead of relying on sequence

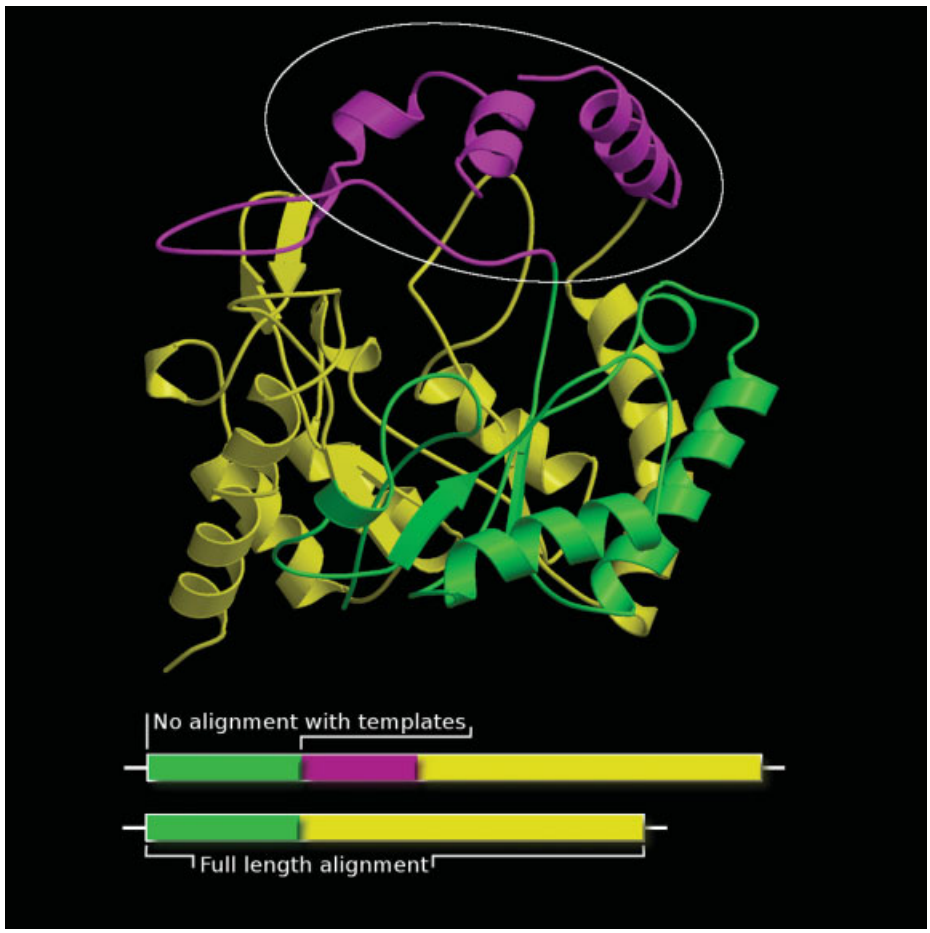

Fig. 1. The "unalignability" problem in T0208. The cartoon represents 3D structure of T0208 with its N-terminal part colored green, C-terminus colored yellow, and the long insertion shown in magenta. Once this insertion was deleted (scheme below), the N-terminal region became readily alignable. This and other structural figures were prepared using combination of Molscript[14] and Raster3D.[15]

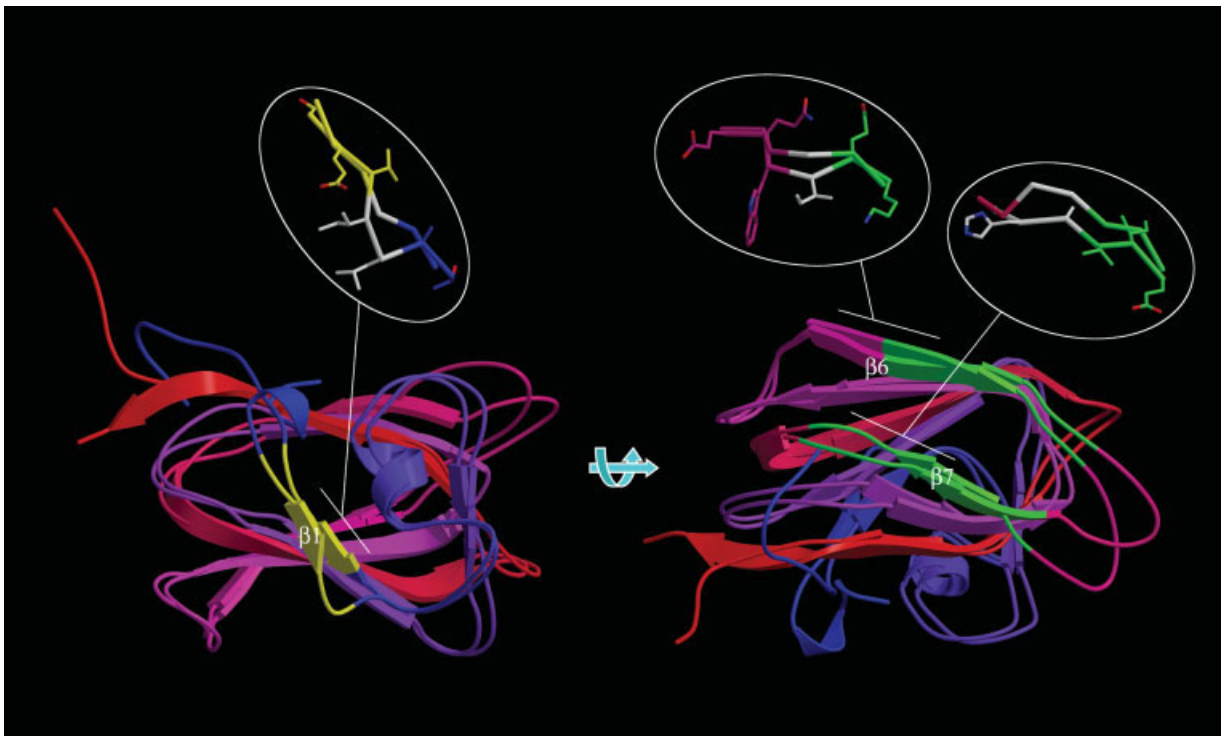

Fig. 2. The local structural features of T0211 made the construction of alignment challenging. The superimposed structures of the target and our model are shown in cartoon representation in opposite orientations. The coloring follows the progression of protein chain from N- (blue) to C-terminus (red). Yellow indicates the region where we made an alignment error because of undetected β-bulge (inset). Regions colored in green are those aligned correctly in our model, but misaligned in majority of other models. The corresponding insets illustrate variable structural features associated with these regions: β-bulge and the place of insertion.

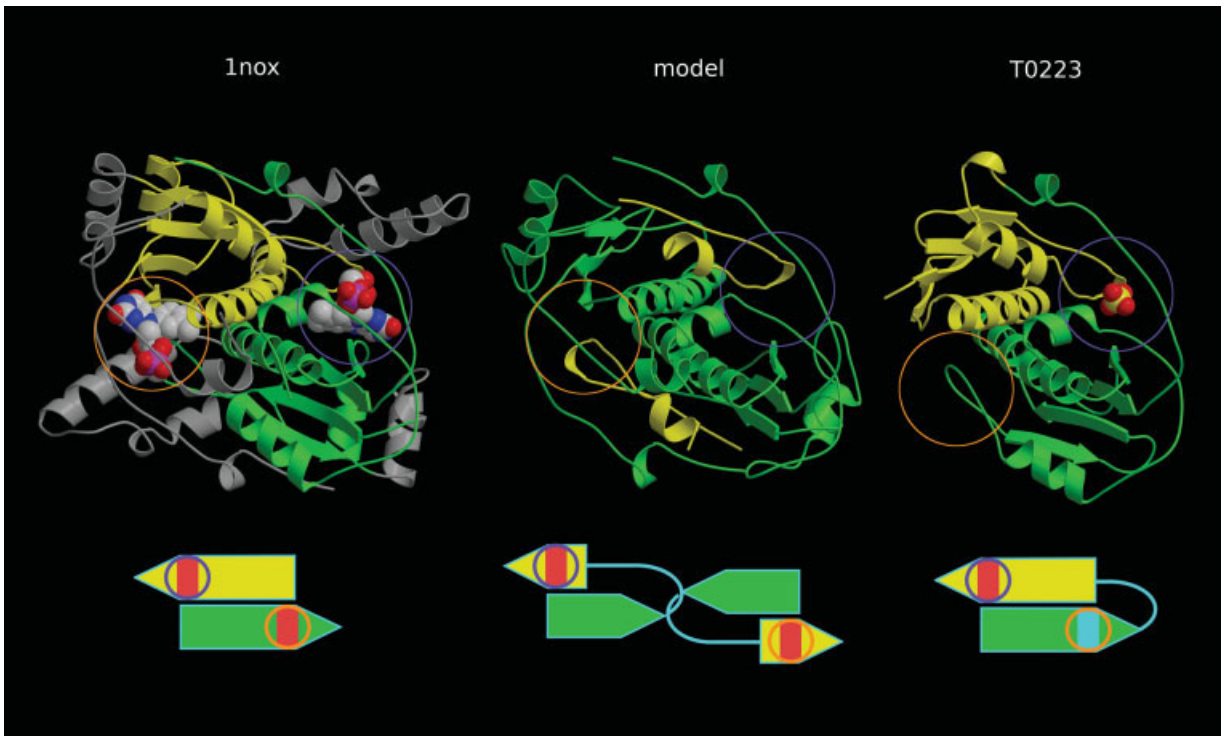

Fig. 3. A deceiving structural organization of T0223. Shown are the dimers of one of the closest structural templates (1nox, left) and our model (middle). The target (right) is a monomer. Yellow and green represents different chains in the template and different domains in the target and model. Binding sites in the template and corresponding regions in the target and model are circled. Below is the schematic representation of chains and their connectivity in the three structures. Apparently inactive binding site in T0223 is colored cyan.

methods to supply the correct alignment we assessed a number of potential alignment variants in the context of the 3D structure. Based on consensus energy evaluation using ProsaII and Verify3D coupled with visual assessment we have correctly identified the presence of β-bulge in β6 and selected the correct length of the β7–β8 connector producing correct alignment for β6 and β7. Yet, we failed to detect the presence of a β-bulge within β1, misaligning its N-terminal part. Surprisingly, even with that error, the fraction of correctly aligned residues in our model is significantly larger than in any other T0211 model. Apparently, the explanation is simple: no other group has correctly aligned both β6 and β7.

Example of an Incorrect Hypothesis Leading to a Wrong Model (T0223)

T0223 is yet another structural genomics target, a putative nitroreductase (PDB code: 1vkw). This target is a good example for illustrating a potential value of human input, because to properly model the complete protein chain it was necessary to use a dimeric template instead of a monomer.

Using a sequence search we have easily established homology between T0223 and flavin-dependent oxireductases. These enzymes are structurally organized as tight dimers, with both chains contributing to each of the two flavin cofactor-binding sites (Fig. 3). Yet, in the T0223 N-terminal region matching structural templates, the characteristic signature of the flavin cofactor-binding site seemed missing. Instead, we have detected such a motif close to the C-terminus and outside the region matching the structural domain. The often observed principle of protein structure conservation led us to generate hypothesis that the target similarly is a dimer with two binding sites. We presumed that the binding motif in T0223 has undergone duplication so that newly created binding motifs replaced the original ones by swapping. We speculated that the first motif might have been lost so that the dimer would still have two binding sites. The length of the target sequence seemed to support this partial duplication hypothesis. The total length of the target is 206 residues, which is shorter than a monomer of most structural templates. Approximately 125 residues of the target could be matched to any individual template, leaving only about 80 C-terminal residues of T0223 unaligned. It seemed logical that the C-terminal region is only a fragment of a structural domain. However, it turns out we were wrong, and the whole structural domain has undergone the duplication. The determined target structure is a monomer containing two stripped down structural domains arranged in exactly the same manner as within the dimeric templates. The first binding site apparently has degenerated, and only the second site remains functional, judging by the bound sulfate ion, corresponding to the phosphate of flavin mononucleotide. Thus, although we produced a model that looks similar to the target structure the hypothetical two-chain architecture is not the correct answer. For this target many automatic servers modeled the first domain

comparably to human groups. In contrast, none of the automatic servers produced any reasonable model for the second domain. Not many human groups were able to correctly infer evolutionary history and produce satisfactory models for the whole protein either. However, those few who did outperformed servers by a large margin (20–30% GDT_TS). Modeling of this target taught us a lesson that human interpretation of the biological data sometimes can be very helpful in modeling, but care should be taken in interpreting the data.

CONCLUSIONS

Our CASP6 results have shown that in template-based modeling the use of multiple templates is often beneficial. However, with the increasing number of available templates a simplistic combination of template structures is perhaps not the best way forward. Combining available structural information at the level of local template fragments may be more effective in increasing the accuracy of models. Yet the ability to select the most accurate structure will require more effective energy-based methods. The same applies to sequence–structure alignments. Our results and results of other groups in CASP6 suggest that in most cases choosing the correct alignment is not a problem of insufficient sampling of different alignment variants. Alignment errors seem to seep in because of an inability to distinguish the correct alignment variant from the incorrect one. Along with the development of better methods to assess alignments within the context of 3D structure it seems important to be able to "defrost" the inherited template backbone. Some of the correct alignments simply do not look plausible in the context of the template scaffold because of backbone deviation that perhaps could be reduced during simultaneous relaxation/refinement of the local structure. In CASP6, the performance of automatic methods has increased substantially, and in high homology modeling human input is often detrimental. However, in modeling distantly related proteins, especially those having unusual structures, human input may be of critical importance.

ACKNOWLEDGMENTS

We thank CASP6 organizers and assessors for their hard work, and experimentalists for providing their protein structures.

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. Venclovas Č. Comparative modeling of CASP4 target proteins: combining results of sequence search with three-dimensional structure assessment. Proteins 2001;Suppl 5:47–54.
3. Venclovas Č. Comparative modeling in CASP5: progress is evident, but alignment errors remain a significant hindrance. Proteins 2003;53(Suppl 6):380–388.
4. 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.
5. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 2003;19:1015–1018.
6. Kurowski MA, Bujnicki JM. GeneSilico protein structure prediction meta-server. Nucleic Acids Res 2003;31:3305–3307.
7. Holm L, Sander C. Mapping the protein universe. Science 1996;273:595–603.
8. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. Proteins 1993;17:355–362.
9. Luthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. Nature 1992;356:83–85.
10. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
11. Canutescu AA, Shelenkov AA, Dunbrack RL Jr. A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci 2003;12:2001–2014.
12. Zemla A. LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res 2003;31:3370–3374.
13. Zemla A, Venclovas Č, Moult J, Fidelis K. Processing and analysis of CASP3 protein structure predictions. Proteins 1999;Suppl 3:22–29.
14. Kraulis PJ. Molscript—a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:946–950.
15. Merritt EA, Bacon DJ. Raster3D: photorealistic molecular graphics. Methods Enzymol 1997;277:505–524.