Protein structure prediction in CASP6 using CHIMERA and FAMS

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · JANUARY 2005

Impact Factor: 2.63 · DOI: 10.1002/prot.20728 · Source: PubMed

6 AUTHORS, INCLUDING:

Genki Terashi
Kitasato University
24 PUBLICATIONS **203** CITATIONS
SEE PROFILE

Kazuhiko Kanou
National Institute of Infectious Diseases, To…
34 PUBLICATIONS **192** CITATIONS
SEE PROFILE

Hideaki Umeyama
Kitasato University
165 PUBLICATIONS **2,533** CITATIONS
SEE PROFILE

Protein Structure Prediction in CASP6 Using CHIMERA and FAMS

Mayuko Takeda-Shitaka,* Genki Terashi, Daisuke Takaya, Kazuhiko Kanou, Mitsuo Iwadate, and Hideaki Umeyama

School of Pharmaceutical Sciences, Kitasato University, Tokyo, Japan

ABSTRACT **In CASP6, the *CHIMERA-group* predicted full-atom models of all targets using SKE-CHIMERA, a Web-user interface system for protein structure prediction that allows human intervention at necessary stages; we used a lot of information from our own data and from publicly available data. Using SKE-CHIMERA, we iterated manual step (template selection and alignment by the in-house program CHIMERA) and automatic step (three-dimensional model building by the in-house program FAMS). The official CASP6 assessment showed that *CHIMERA-group* was one of the most successful predictors in homology modeling, especially for FR/H (Fold Recognition/Homologous). In this article, we introduce the method of *CHIMERA-group* and discuss its successes and failures in CASP6. Proteins 2005;Suppl 7:122–127.**

Key words: fold recognition; homology modeling; protein structure prediction; CHIMERA; FAMS; SKE-CHIMERA; human intervention; CASP

INTRODUCTION

Genome sequencing projects have generated huge amounts of protein sequence information. Moreover, the structural genomics project that aims to determine experimentally at least one representative three-dimensional (3D) structure from every protein family has started. This project seeks to determine approximately 10,000 structures over a period of 5–10 years. Based on these representative structures, 3D structures of the majority of the proteins encoded in genomes might be predicted by homology modeling. For this reason, homology modeling has become a key component in the postgenomic era. Our laboratory developed the CHIMERA modeling system,[1–3] which predicts protein structures based on homology modeling methods using more than one reference protein. CHIMERA is an interactive homology modeling system that enables human intervention at necessary stages (template search, alignment, and 3D structure construction), which improves the model quality. After development of CHIMERA, our laboratory also developed FAMS (a fully automated homology modeling system)[3–6] by automating procedures for 3D structure construction in CHIMERA. FAMS only requires an alignment as input and constructs 3D model structures automatically. Fully automated approaches like FAMS are essential for large-scale genome modeling. In CASP6, we entered three groups (*FAMD-server* and *FAMS-server* as servers, and *CHIMERA-group* as human predictor; note that in this article we use italics to refer to group names in order to distinguish them from software names). In all three groups, we used FAMS for 3D structure construction. On the other hand, the methods of *CHIMERA-group* and of the two servers were different in template selection and alignment steps. *FAMD-* and *FAMS-servers* used alignments obtained automatically by standard sequence analysis methods such as PSI-BLAST. In many cases, however, automatic alignments included errors. Therefore, *CHIMERA-group* selected templates and generated alignments with human intervention using our modeling system CHIMERA. The official CASP6 assessment showed that *CHIMERA-group* was one of the most successful predictors in homology modeling (http://predictioncenter.org/casp6/meeting/presentations/CASP6_Program.doc). Here, we describe the methods of *CHIMERA-group* and discuss the successes and failures of our predictions.

MATERIALS AND METHODS

SKE-CHIMERA: A Web-User Interface System for Protein Structure Prediction

In order to construct high-quality models, a lot of information necessary for protein structure prediction must be considered. Moreover, in many cases, human intervention improves the quality of models. We developed SKE-CHIMERA, a Web-user interface system for protein structure prediction, through which a lot of data we prepare can be analyzed, and homology modeling is easily carried out with human intervention at necessary stages. In CASP6, *CHIMERA-group* used SKE-CHIMERA for all targets. As described in the Introduction, *CHIMERA-group* selected templates and generated alignments with human intervention using CHIMERA, and constructed 3D structures automatically using FAMS. Steps using CHIMERA and FAMS were iterated using SKE-CHIMERA, until the final model structure was built. In the following

*Correspondence to: Mayuko Takeda-Shitaka, School of Pharmaceutical Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan. E-mail: shitakam@pharm.kitasato-u.ac.jp

Received 15 April 2005; Accepted 21 June 2005

Published online 26 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20728

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

section, we introduce methods of *CHIMERA-group* using SKE-CHIMERA.

Automatic Preparation of Initial Sequence–Template Alignments

Initial alignment variants between each target and its corresponding templates were generated using as many programs as possible, because different programs gave various candidates of templates and alignments. We used six standard sequence search methods: BLAST,[7] PSI-BLAST,[8] RPS-BLAST,[9] IMPALA,[10] FASTA,[11] and HMM-BLAST (http://hmmer.wustl.edu/). In addition to these six programs, we used two programs, PSF-BLAST and PRED-FASTA, that were previously developed in our laboratory. PSF-BLAST is a variant of PSI-BLAST algorithm, and PRED-FASTA was developed based on the algorithms of PSIPRED[12] and FASTA. All alignments derived from these eight methods were ranked by our *FAMD-* and *FAMS-servers* using our original scoring functions that attempted to select the alignment whose resulting 3D structure had the highest Global Distance Test Total Score (GDT_TS).[13] Our scoring functions took into account alignment length, sequence identity, and degree of agreement between secondary structure of the template assigned by STRIDE[14] and secondary structure prediction for target using PSIPRED. To prepare initial alignment variants further, publicly available prediction data for human predictors in CASP6 (alignments submitted in AL format by CAFASP4 servers) were used. Although 3D atomic coordinates that were submitted in tertiary structure (TS) format by CAFASP4 servers were also publicly available, we did not use them, because we wanted to evaluate the ability of our original software FAMS to construct 3D structures.

Selection of Templates

Since a certain method cannot always give the best solutions, we equally used all initial alignments for template selection and alignment. All initial alignments derived from the eight programs and CAFASP4 servers were observed through SKE-CHIMERA. In addition, sequence identity, *e*-value, secondary structure predictions using PSIPRED, SAM-T02-DSSP, SAM-T02-STRIDE, ROBETTA-JUFO-3D, and PROFSEC (http://www.cs.bgu.ac.il/~dfischer/CAFASP4/), scores of *FAMD-* and *FAMS-servers*, and conservation of structurally and functionally important residues were visible through SKE-CHIMERA. We selected several templates from a number of candidates for each target, taking into account the above data. In addition, visual inspection of 3D structures of template candidates and literature searches were useful. In some cases, 3D jury scores shown in the CAFASP4 website were useful. The final selection was made after constructing and evaluating 3D models.

Generating Our Own Alignments

We defined reliable regions and questionable regions in the alignments by comparing all initial alignments using SKE-CHIMERA. Then, we manually generated our own alignments with emphasis on the questionable regions, using the following information:

1. Secondary structure predictions, as mentioned above.
2. 3D structures of candidate templates and their homologs were superimposed by the Combinatorial Extension (CE) algorithm.[15] Multiple sequence alignment of homologs and visual inspection of superimposed homologs gave a good indication of exact positions of gaps in insertion/deletion regions.
3. The sequence of each target protein was compared with all other sequences in the nonredundant protein sequence database (nr), provided by the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) using BLAST. Multiple sequence alignment consisting of up to 50 sequences (sequence identity $>$ 30% and *e*-value $<$ 0.001) from nr was useful for refinement of insertion/deletion regions (and template selection).
4. Literature searches were made to obtain available biological information for any of the proteins in the same family as the target.

Automatic Construction of 3D Structures

All alignments generated manually were evaluated by constructing full-atom 3D structures using FAMS. Because FAMS requires only an alignment as input and constructs a 3D model structure automatically, we could construct a lot of models based on various alignments for each target using PC clusters. Procedures for FAMS have been reported in details in a previous article from our laboratory.[5] There are three steps in FAMS: (1) construction of Cα atoms, (2) construction of main-chain atoms, and (3) construction of side-chain atoms with main-chain optimization. Main-chain atoms are constructed based on similarities between environmental residues at topologically equivalent positions in the template (local space homology). Side-chain atoms are constructed by iterative cycles of side-chain generation and main-chain optimization. Side-chain generation is based on conservation of side-chain conformation for each residue within homologous proteins.[16] Main-chain atoms are optimized with the fixed side-chain conformations. In CASP6, the loop search process of FAMS was modified to select fragments from proteins of the same family as templates. Therefore, selection of fragments was based on root-mean-square deviation (RMSD) of fitting and degree of SCOP[17] ID agreement between template and fragment. Although FAMS can construct model structures using multiple templates, most of the models were constructed using a single, main template in CASP6 because of limited time.

Evaluation of Model Structures

Models were evaluated by in-house programs that checked for unfavorable contacts between atoms, unnatural chiral centers, *cis*-peptides except *cis*-proline, hydrophobic interaction, and continuity of secondary structure elements. Main-chain ϕ–ψ angles were evaluated by Ramachandran plots made by PROCHECK.[18] Secondary struc-

tures of the models assigned by STRIDE were compared to the five secondary structure predictions. We analyzed all these evaluation results using SKE-CHIMERA. Model structures were also checked by visual inspection (for exposed hydrophobic residues, buried charged residues, conservation of specific residues within the family, etc.). If necessary, template selection, alignment refinement, model building, and evaluation were iterated using SKE-CHIMERA. Finally, we manually selected the best model for the target among a number of models according to evaluations described above. We submitted one model per target, because the general rule on CASP6 homepage was that most of the evaluation and assessment will focus on the model labeled "1."

RESULTS

Because domain definition and categorization of each target by CASP6 assessors were unknown during the CASP6 experiment, we submitted models of all CASP6 targets no matter what their level of homology to proteins of known structures, while we aimed to predict homology modeling targets. In the official CASP6 assessment, targets were classified into five categories: homology modeling [Comparative Modeling (CM)/Easy (25 domains), CM/Hard (18 domains), Fold Recognition/Homologous (FR/H) (22 domains)], and nonhomology modeling [FR/A (Analogous) (15 domains) and New Fold (NF) (10 domains)]. In homology modeling, CM/Easy was the easiest category and FR/H was the most difficult category. According to the official CASP6 assessment, *CHIMERA-group* succeeded in the homology modeling category. Especially for the FR/H category, *CHIMERA-group* did quite well. Automated assessment of CASP6 models using the Template Modeling Score (TM-score)[19] also showed similar results (http://www.bioinformatics.buffalo.edu/new_buffalo/people/zhang6/casp6/). In the following, we describe results of *CHIMERA-group* for homology modeling.

CM/Easy and CM/Hard

According to the CASP6 official assessment (http://predictioncenter.org/casp6/abstracts/CASP6_Tables_light.txt), almost all of our models in CM/Easy and CM/Hard categories were of higher quality than those of other groups. In these categories, automatic template selections by PSI-BLAST and its relatives were accurate in many cases, so that the quality of the models largely depended on the alignment step. Therefore, manual refinement of alignments may be the major reason for our high-quality models. For example, for T0229_2, detailed refinement of gap placement in the insertion/deletion regions using SKE-CHIMERA resulted in the highest GDT_TS among the participants. Only in three domains (T0233_1, T0280_1, and T0232_2) out of 43 domains were GDT_TS scores of *CHIMERA-group* less than average. Therefore, we analyzed the failures of these three domains.

T0233, anthranilate phosphoribosyltransferase (AnPRT) 2 from *Nostoc* sp. Pcc 7120, was a CM/Easy target. We selected three experimental structures of AnPRTs from *Sulfolobus solfataricus* [Protein Data Bank[20] (PDB) ID: 1o17), from *Thermus thermophilus* (1v8g), and from *Pectobacterium carotovorum* (1kgz) as templates, and constructed three models based on these three templates, respectively. Because sequence identity between the target and 1v8g was the highest (40.1%), we submitted the model based on 1v8g out of the three models. In CASP6 official assessment, T0233 was divided into two domains (T0233_1 and T0233_2) that were evaluated separately. GDT_TS scores of *CHIMERA-group* were 79.55 and 84.72 for T0233_1 and T0233_2, respectively. Comparison between the native structure of the target (1vqu) and our three models showed that the best models for T0233_1 and T0233_2 were based on 1o17 (GDT_TS = 92.80) and 1v8g (GDT_TS = 84.72), respectively. Moreover, 1o17 and 1v8g showed the highest sequence identities in the regions of T0233_1 (28.8 %) and T0233_2 (43.4 %), respectively. If we had divided this target into two domains and used the best template for T0233_1, we could have constructed a more accurate model for T0233_1.

T0280 was TT1426 from *T. thermophilus* HB8 (hypothetical protein with a predicted phosphoribosyltransferase domain). The native structure of T0280 (1wd5) resembles phosphoribosylpyrophosphate synthetase from *Bacillus subtilis* (1dkr). However, the significant difference between them is that 1wd5 has a large inserted domain in comparison with 1dkr. Therefore, T0280 was divided into two domains in the CASP6 official assessment: T0280_1 that corresponded to 1dkr (residues 5–52 and 115–179; CM/easy), and T0280_2 that corresponded to the inserted domain (residues 53–103; FR/A). For the C-terminal half of T0280_1 (residues 115–179), we could correctly select 1dkr as template, and could construct a model structure. However, we could not construct the N-terminal half of T0280_1 (residues 5–52) because of alignment errors caused by failure to predict the inserted domain. After the native structure was given, we found that the N-terminal half of T0280_1 was aligned with 1dkr by IMPALA in SKE-CHIMERA, although the score calculated using our original scoring function was low. In our system, the N-terminal region (with low score) and the C-terminal region (with high score) of T0280_1 were aligned with 1dkr separately, and we failed to combine them. A reliable method for dividing target sequences into domains (or predicting large insertion) is necessary for better models.

T0232, Atu5508 from *Agrobacterium tumefaciens*, was a CM/Hard target. Because the native structure of this target is not available to the public at present, details cannot be discussed in this article. In short, the failures for T0232_2, which was an α-helix-rich domain, were due to alignment errors of α-helices.

FR/H

For difficult targets (low homology cases), it is hard to find templates by using standard sequence search methods such as PSI-BLAST. We also used FR-based alignments submitted in AL (Alignment) format by CAFASP4 servers that were publicly available for human predictors in CASP6. In the FR/H category, template selection was

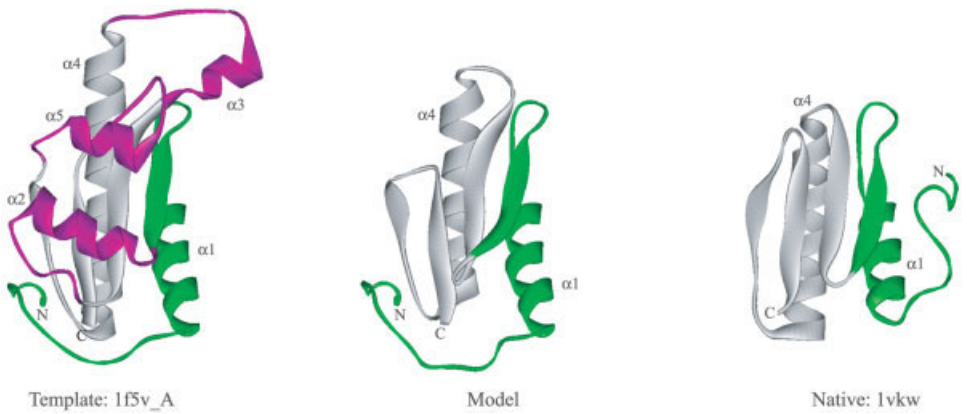

Fig. 1. Modeling of T0223_2 [putative nitroreductase (Tm1586) from *T. maritima*]. Template structure (1f5v_A), our model, and the native structure of T0223_2 (1vkw) are shown. The N-terminal region that was aligned automatically is shown in green. The C-terminal region that was aligned with human intervention is shown in white and magenta. We predicted that the three α-helices of 1f5v_A, shown in magenta (α2, α3, and α5), were missed in the target.

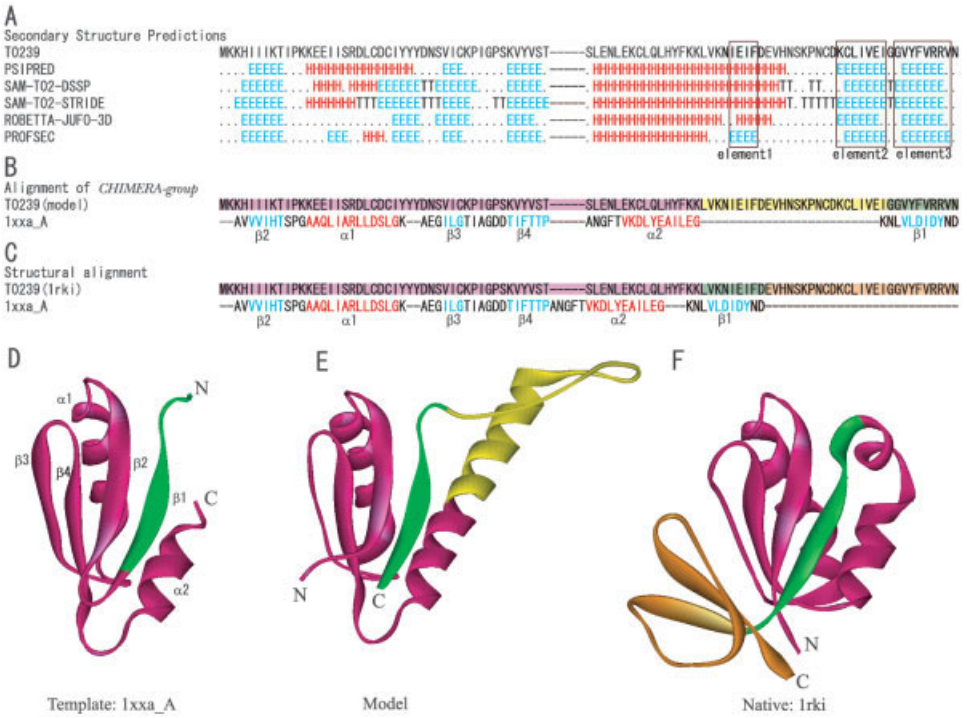

Fig. 2. Modeling of T0239 (hypothetical protein PAE0736 from *P. aerophilum*). (**A**) Five secondary structure predictions for T0239. H denotes α-helix and E denotes β-strand. (**B**) Alignment between T0239 and 1xxa_A by *CHIMERA-group*. Regions highlighted in magenta, yellow, and green in T0239 (model) correspond to the regions shown in the same colors in (E). α-helices and β-strands of 1xxa_A are shown in red and blue, respectively. (**C**) Structural alignment between the native structure (1rki) and the template structure (1xxa_A). Regions highlighted in magenta, green, and orange in T0239 (1rki) correspond to regions shown in the same colors in (F). α-helices and β-strands of 1xxa_A are shown in red and blue, respectively. (**D**) Structure of the template (1xxa_A). (**E**) Structure of our model. (**F**) Structure of the native protein (1rki).

not easy, even if fold recognition data from CAFASP4 servers were used, because different servers showed different candidate templates in many cases. Therefore, correctness of template selection was a significant determinant for the quality of the model. According to the official CASP6 assessment, predictions of *CHIMERA-group* for T0214, T0223_2, T0239 (this target was FR/A), T0206, T0249_1, T0227, T0262_2, and T0243 were particularly successful. In the following paragraph, we illustrate three predictions with a high *Z*-score of GDT_TS from our predictions.

T0223_2, a putative nitroreductase (Tm1586) from *Thermotoga maritima*, is an example showing the necessity of manual alignment. Although the sequence identity was

low (15%), the template (1f5v_A) was automatically found using FR servers. However, only a small portion in the N-terminal region of T0233_2 was aligned with 1f5v_A (Fig. 1, green) by the servers. Therefore, we had to manually align the C-terminal region. In the C-terminal region, the target sequence was much shorter than 1f5v_A by 60 residues; therefore, we predicted that three α-helices in 1f5v_A (α2, α3, and α5 in Fig. 1) were missed in the target based on secondary structure predictions and visual inspection of template structure. The native structure of the target (1vkw) showed that our prediction was very successful (Fig. 1), although the α4 helix was misaligned by several residues.

In the case of T0239, a hypothetical protein PAE0736 from *Pyrobaculum aerophilum*, different FR servers showed different templates, which meant that template selection was difficult. Therefore, we manually searched for the best template among a lot of candidates, based on the secondary structure predictions and visual inspection of their structures. We selected the C-terminal domain of arginine repressor from *Escherichia coli* (1xxa_A) as a template. Because sequence identity was very low (about 10%), and there was a large insertion in the target (about 30 residues in total), alignment was also difficult. The N-terminal half of the target was correctly aligned with 1xxa_A [shown in magenta in Fig. 2(B, D, and E)] because secondary structure predictions [Fig. 2(A)] and the secondary structure of 1xxa_A agreed well. In the C-terminal region, secondary structure predictions suggested that there were three elements [elements 1–3 in Fig. 2(A)]. By visual inspection of 1xxa_A, we predicted that element 3 of the target corresponded to the N-terminal β-strand of 1xxa_A [β1; shown in green in Fig. 2(B, D, and E)], and that elements 1 and 2 were part of a large insertion [Fig. 2(B and E), yellow]. However, the native structure of the target (1rki) showed that element 1 of the target corresponded to β1 of 1xxa_A, and that elements 2 and 3 were part of a large insertion [Fig. 2(C and F), orange]. We were misled by the fact that four of five secondary structure predictions suggested that element 1 was an α-helix, not a β-strand. For this target, we obtained the high *Z*-score because template selection was correct, but further improvements were needed in the C-terminal region. According to the results of CASP6, no group could predict the large insertion in the C-terminal region.

T0214 and T0227 are good examples for the importance of manual template selection and manual alignment. For these targets, almost all of the FR servers failed to select the correct template. The correct template for both targets was hypothetical protein (Gi:29377587) from *Enterococcus faecalis* v583 (1t62). We manually found this template by comparing target sequences with all candidate templates mainly based on secondary structure predictions, visual inspection of template structures, and publicly available discussions on FORCASP site of CASP6 website. Moreover, we could manually generate better alignments compared to other groups. Successful manual tasks resulted in high *Z*-scores of these targets.

DISCUSSION

CASP6 results suggest that our method including human intervention using SKE-CHIMERA was successful in homology modeling. Automatic prediction data were valuable starting points for our predictions. Using SKE-CHIMERA, we could find errors in automatic prediction data and correct them with human intervention using our own data and publicly available information. Also the ability of FAMS to construct 3D models accurately was one factor for success. In CASP6, the ranking and *Z*-score of *CHIMERA-group* in FR/H was slightly higher than those in CM/Easy and CM/Hard. Among the three categories, FR/H was the most difficult, and there were many errors in template selection and alignment by automatic servers; therefore, human intervention through SKE-CHIMERA was probably most effective in FR/H.

Although *CHIMERA-group* succeeded in CASP6, further improvements are required in our method. First, our alignments were not perfect; therefore, we must improve alignment methods. Second, model evaluation must be improved, because we failed to select the best model among the constructed models for many targets. Third, accurate prediction of domain classification must be done. Fourth, FAMS constructs structural variable regions by loop searches in the PDB, but in order to construct finer structures, an ab initio protocol must be included in FAMS procedures. Fifth, in CASP6, we used multiple templates in the manual template selection and alignment steps by CHIMERA. However, we used a single, main template in the 3D structure construction step by FAMS because of limited time, though FAMS can construct models using multiple templates. We should have used multiple templates also in the 3D structure construction step, because the use of multiple templates in the model construction step can significantly improve model accuracy. Last, many of the steps done with human intervention are worth automating and are automatable. Therefore, we will automate our procedures and improve SKE-CHIMERA to become a more facile system for human predictors. In CASP6, *CHIMERA-group* resulted in a better performance than *FAMD-* and *FAMS-servers* in all categories. We will also improve *FAMD-* and *FAMS-servers* by incorporating human procedures of *CHIMERA-group*.

ACKNOWLEDGMENTS

Our thanks to the CASP6 organizers and assessors, and structural biologists who supplied targets for CASP6, and to Chieko Chiba and Hirokazu Tanaka for their helpful discussions.

REFERENCES

1. Yoneda T, Komooka H, Umeyama H. A computer modeling study of the interaction between tissue factor pathway inhibitor and blood coagulation factor Xa. J Protein Chem 1997;16:597–605.
2. Takeda-Shitaka M, Umeyama H. Elucidation of the cause for reduced activity of abnormal human plasmin containing an Ala55–Thr mutation: importance of highly conserved Ala55 in serine proteases. FEBS Lett 1998;425:448–452.
3. Takeda-Shitaka M, Takaya D, Chiba C, Tanaka H, Umeyama H. Protein structure prediction in structure based drug design. Curr Med Chem 2004;11:551–558.

4. Takeda-Shitaka M, Nojima H, Takaya D, Kanou K, Iwadate M, Umeyama H. Evaluation of homology modeling of the SARS coronavirus main protease for structure based drug design. Chem Pharm Bull 2004;52:643–645.
5. Ogata K, Umeyama H. An automatic homology modeling method consisting of database searches and simulated annealing. J Mol Graph Model 2000;18:258–272.
6. wadate M, Ebisawa K, Umeyama H. Comparative modeling of CAFASP2 competition. Chem-Bio Informatics J 2001;4:136–148.
7. Altchul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.
8. Altchul SF, Madden TL, Schäffer 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.
9. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res 2002;30:281–283.
10. Schäffer AA, Wolf YI, Ponting CP, Koonin EV, Aravind L, Altchul SF. IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. Bioinformatics 1999;15:1000–1011.
11. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 1988;85:2444–2448.
12. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.
13. Zemla A, Venclovas C, Moult J, Fidelis K. Processing and analysis of CASP3 protein structure predictions. Proteins 1999;Suppl 3:22–29.
14. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. Proteins 1995;23:566–579.
15. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 1998;11:739–747.
16. Ogata K, Umeyama H. The role played by environmental residues on sidechain torsional angles within homologous families of proteins: a new method of sidechain modeling. Proteins 1998;31: 355–369.
17. Lo Conte L, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Res 2002;30:264–267.
18. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
19. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins 2004;57:702–710.
20. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.