PROTEINS: Structure, Function, and Bioinformatics Suppl 7:157–166 (2005)

Prediction of CASP6 Structures Using Automated Robetta Protocols

Dylan Chivian,[1†] David E. Kim,[1†] Lars Malmström,[1] Jack Schonbrun,[2] Carol A. Rohl,[3] and David Baker[1,2*]
[1]*Department of Biochemistry, University of Washington, Seattle, Washington*
[2]*Howard Hughes Medical Institute, Seattle, Washington*
[3]*Department of Biomolecular Engineering, University of California, Santa Cruz, California*

ABSTRACT **The Robetta server and revised automatic protocols were used to predict structures for CASP6 targets. Robetta is a publicly available protein structure prediction server (http://robetta.bakerlab.org/) that uses the Rosetta de novo and homology modeling structure prediction methods. We incorporated some of the lessons learned in the CASP5 experiment into the server prior to participating in CASP6. We additionally tested new ideas that were amenable to full-automation with an eye toward improving the server. We find that the Robetta server shows the greatest promise for the more challenging targets. The most significant finding from CASP5, that automated protocols can be roughly comparable in ability with the better human-intervention predictors, is repeated here in CASP6. Proteins 2005;Suppl 7:157–166.**

Key words: Rosetta; fragment assembly; de novo modeling; homology modeling; parametric alignment ensemble

INTRODUCTION

Full automation of protein structure prediction is a desirable goal as it opens the door to genome-level protein structure modeling and, equally importantly, provides a stringent test of the principles underlying prediction methods unadulterated by the powerful influence of human intuition. Automated methods were for the first time found in CASP5 to be approaching the abilities of the best human modelers.[1,2] However, as successful innovations are introduced to the field, even the best automated methods must keep pace in order to maintain their relevance. Automation and incorporation of successful concepts from human-applied methodologies is therefore of paramount importance.

The publicly available Robetta server (http://robetta.bakerlab.org/) was found in CASP5 to be among the better methods for protein structure prediction,[1–3] comparing favorably with both fully automated methods and those that allow for human intervention. Robetta, which has been previously described,[4,5] employs the principle that a protein chain should be modeled at the domain level following the approach that is most appropriate for each domain. The best method for predicting the structure of a protein depends on whether it has sequence homology to a protein of known structure. If a detectable similarity exists, then a relatively accurate model can be built using the known structure as a template. In the absence of such similarity, models can be built using de novo prediction methods, which do not rely on a template structure. In many cases, hybrid template-based/de novo methods may be most appropriate: portions of a given target may be modeled based on a template, while it may only be possible to model long variable loops or extra domains or extensions not contained in the template using de novo methods. The Robetta server attempts to provide the best possible model for the entire length of the protein chain by combining template-based and de novo protocols.

METHODS

Robetta uses the Rosetta fragment-assembly technique[6,7] to build models of protein domains in both template-based and de novo modes. Modeling is performed at the domain level based on the assumption that domains are autonomously folding units.[8–10] Since protein chains are often comprised of more than one domain, it is essential that any server that attempts to model the full length of a query in domain-sized pieces determine the location of putative domains, assign each of those domains to the appropriate template-based or de novo protocol, and ideally restore chain connectivity between the domains by assembling the domain models into a single multi-domain prediction.

As has been described previously, the initial step, called "Ginzu"[4] (and see accompanying report in this issue), involves screening the query sequence for regions that possess a homolog with an experimentally characterized structure with BLAST, PSI-BLAST,[11,12] FFAS03,[13,14] and 3D-Jury,[15] followed by cutting the sequence into putative domains based on matches to Pfam sequence families,[16] multiple sequence information, and predicted secondary structure information. Any detected parents and the re-

---

The Supplementary Material referred to in this article can be found online at www.interscience.wiley.com/jpages/0887-3585/suppmat/

†Dylan Chivian and David E. Kim contributed equally to this paper.

*Correspondence to: David Baker, Dept. of Biochemistry and HHMI, University of Washington, Box 357350, Seattle, Washington 98195. E-mail: dabaker@u.washington.edu

Received 15 April 2005; Accepted 24 June 2005

Published online 26 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20733

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

gions of the query with which they are associated are stored and assigned to the template-based modeling protocol. Putative domain regions without corresponding parents are modeled by the Rosetta de novo protocol.[17,18]

After domain parsing, each putative domain then follows its assigned protocol track. For the domains to be modeled de novo, Robetta employed a protocol that was quite similar to that used in CASP5, which generates large numbers of alternate "decoy" conformations for the target and up to two sequence homologs, and subsequently filters the decoy ensemble to remove nonprotein-like conformations and clusters the remaining structures to identify broad low free energy minima.[17] However, as we reported previously,[4] we repaired a bug in the Rosetta energy function that contributed to a reduction in quality of the decoys in the ensemble. Also as reported, shortly after CASP5 we revised the decoy clustering method to use a dynamic RMSD100[19] threshold, which varies depending on the length of the target, rather than a fixed RMSD threshold for determining membership in the same cluster, thus allowing for more diversity between the top clusters, from which the final models are selected. We additionally have increased the number of decoys generated (from 4,000 to 10,000 for the query and from 2,000 to 5,000 for each of up to two sequence homologs).

For homology modeled domains, Robetta decouples the detection of the parent from the alignment to that parent. It ignores the alignment from the detection method, and instead attempts to obtain a superior alignment using our "K*Sync" alignment method.[4] K*Sync takes into account evolutionary sequence information for both the query and the parent, secondary structure information, and information from multiple structural alignments by our StrAD-Stack method (D.C., manuscript in preparation) to indicate regions that are likely to be structurally obligate to the fold. In an approach employed by our human group in CASP5,[20] Robetta generates a decoy ensemble by first generating a parametric alignment ensemble[21,22] using K*Sync. This is accomplished by varying the weights on pair and gap terms in the dynamic programming alignment, by changing the source of the secondary structure predictions (PSIPRED,[23] SAM-T99,[24] and JUFO[25]), the number of rounds and stringency for inclusion in the sequence family found by PSI-BLAST[11,12] used to obtain the residue substitution profiles, and the stringency of the structures included in the StrAD-Stack multiple structural alignment. From each alignment a template is generated, and variable regions are then modeled with a version of the Rosetta de novo method that allows the conformations of variable regions to be sampled in the context of a fixed template.[26] Final models were then selected from this decoy ensemble using low or high resolution energy functions,[27] an approach similar to other groups.[28–30] The first model was the one possessing both reasonable hydrophobic sequestering and a good energy in an all-atom representation (with an attenuated Lennard-Jones repulsive term to allow for problems in the model resulting from the frozen backbone in the template region). The second model submitted was generated from the single default K*Sync alignment. Models 3 through 5 were additional low-energy models from the ensemble that differed from already submitted models.

If a target possesses more than one domain, the separate domain models are then combined into one full-length model by fragment replacement in the putative linker region in order to provide chain connectivity and to attempt to predict the inter-domain interactions. The last step consists of repacking the side chains using a backbone-dependent rotamer library[31] with a Monte Carlo conformational search procedure.[32]

In order to assess our ability to improve the Robetta protocols, we developed several innovations that we felt might allow for better modeling and applied them during CASP6 as a separate group from the Robetta server. This alternate protocol, which we called "Robetta_04," was essentially fully automated, but not run according to the time constraints placed upon server groups.

The Robetta_04 template-based modeling protocol primarily differed from the standard protocol by using cyclic coordinate descent for loop closure,[33] K*Sync alignment ensembles produced from more than one parent, and full-length backbone refinement in conjunction with side-chain centroid energy[27] optimization for Fold Recognition targets. Selection of models from the ensemble for targets with parents detected by PSI-BLAST was done using the full-atom Rosetta potential[27] (again, with an attenuated Lennard-Jones repulsive), with the five models selected in order by energy. The Fold Recognition targets, expected to deviate more from the native than the Comparative Modeling set and therefore to be less amenable to a full-atom representation, were selected from the ensemble using the same Rosetta side-chain centroid energy[27] that was optimized.

The Robetta_04 de novo protocol was almost identical to the Robetta protocol, with the only variation coming from reranking of cluster centers using a confidence function[18] that included matches of candidate models to experimental structures in addition to the contact-order of the model and its match, and a clustering threshold to measure convergence.

Versions and Parameters

BLAST and PSI-BLAST[11,12] parent detections were done using PSI-BLASTv 2.2.6 starting from BLOSUM62[34] against the pdb_seqres.txt[35] and using the nonredundant sequence database from the NCBI (nr). The iterative detection was done via automatic restart from a checkpoint file against the pdb_seqres.txt after five rounds of profile building against the nr, with an e-value for inclusion of 0.001 or closer.

FFAS03[13,14] detections were retained if their score was − 20.0 or better. Although FFAS03 is usually correct in its detection with a score of − 9.5 or better, we chose to turn to 3D-Jury in this "intermediate" regime to obtain detections corroborated by other fold recognition servers.

The variant of 3D-Jury[15] that Robetta employed was most similar to the "3D-Jury-A1" method, that only uses "first-order" Fold Recognition servers as input. These

input servers were: mGenTHREADER,[36,37] FUGUE-2,[38] 3D-PSSM,[39] BIOINBGU,[40] BASIC[41] (dist), and FFAS03.[13,14] Robetta ignores detections under 30 residues and with a 3D-Jury confidence less than 25.0. A confidence between 25.0 and 50.0 was considered a twilight-zone detection, and modeling was performed by both template-based and de novo methods in this regime.

Ginzu uses PSIPREDv2.01[23] and five rounds against the nr with PSI-BLASTv2.2.6 starting from BLOSUM62, e-value for inclusion and reporting 0.001 or closer.

K*Sync default alignments used PSI-BLASTv2.2.6 with BLOSUM62 for two rounds e-value $\leq 10^{-6}$ against the nr followed by one round e-value ≤ 0.001, secondary structure from PSIPREDv2.01, and structural alignment of the parent with structural homologs from StrAD-Stack multiple structural alignments of domains (z-score ≥ 2.5). StrAD-Stack alignments were accomplished by parsing the parent into domains using Taylor's method[42] and structurally aligning parent domains with a nonredundant domain library using the 3D-Pair structure–structure alignment program.[43]

LGA structure-structure alignment and GDT analysis was done with the LGA server[44] (http://predictioncenter.org/local/lga/lga.html). Comparison of models with the target native structure was done with a sequence-dependent fit at 4 Å (using the options: "-3 -sda -o1 -d_4.0 –lw_7 -atoms:CA"), considering the residues for which there was density in the target native structure.

RESULTS AND DISCUSSION

The modeling regime used by Robetta and Robetta_04 in CASP6 for each domain of each target is shown in Table I. The targets are separated into columns based on the classification of the assessors, and the method used by Ginzu to detect the parent for the domain is indicated next to the target. Robetta processed the targets in a fashion roughly following the classification of the targets by the assessors. The greatest exception to this was for some of the more challenging Fold Recognition targets, for which a parent was not confidently detected. Those with a weak 3D-Jury confidence (see Methods section for thresholds), which we refer to as "twilight-zone" targets, were modeled by both template-based and de novo approaches. The targets with the least confident detections by 3D-Jury we modeled using only the de novo protocol, rather than using a parent with a detection confidence at a level dominated by incorrect predictions. One complication in the analysis results from the application of the Robetta_04 protocol near the date that models were due in contrast to the Robetta server method which was run at the time of release of the targets. This resulted in different methods used by Robetta and Robetta_04 to model the few targets that changed their detection level during the intervening time span (T0216, T0226, T0243).

We leave more thorough discussion of Robetta and Robetta_04 model quality with respect to the field as a whole to the assessors. As our goal was to ascertain the possibilities for improving the Robetta method, we instead compare the quality of models from the Robetta server with those from the Robetta_04 method and with the models submitted by our human group (for a discussion of the human group's methods, see the accompanying report in this issue). In our comparison, we use as a baseline the model that would have been generated had we simply stopped with the model produced by the top confidence (or longest similar confidence) detection, since we must improve upon these models to consider our method as adding value. We therefore have grouped the targets into categories defined by the method that Robetta employed for detection of a parent during CASP6 for this analysis.

Table II shows the performance of methods within each detection regime as measured by GDT_TS. Several methods are shown: the baseline PSI-BLAST, FFAS03, or 3D-Jury method ("Base"), the Robetta server ("Rob"), the Robetta_04 method ("Rob_04"), and our human group ("Baker") (for the target-specific scores used to generate Table II, see Supplementary Table I). The average GDT_TS scores for the first model ("First") and the best of the five submitted models ("Best") are reported. The greatest

TABLE I. Robetta Modeling Protocol and Parent Detection Source[†]

CM/easy	CM/hard	FR/H	FR/A	NF
b T204	p T196	j T197	j T198	d T201
b T229_1	i T199_1	i T199_2	i T199_3	d T209_2
b T229_2	p T200	j T202_1	d T209_1	d[b] T216_1
b T231	f T205	f T203	d T212	d[b] T216_2
b T233_1	p T208	p T206	d T215	t T238
b T233_2	f T211	n T213	t T230	t T241_1
p T235_1	p T222_1	n T214	p T235_2	t T241_2
b T240	p T223_1	f T222_2	d T239	d T242
b T244	i[a] T226_1	p T223_2	t T248_1	t T248_2
b T246	p T232_1	i T224	t T248_3	
b T247_1	p T232_2	d T227	j T262_1	
b T247_3	b T264_2	f T228_1	d T272_1	
b T264_1	p T265	f T228_2	d T272_2	
b T266	p T267	t T237_1	d T273	
b T268_1	b T269_2	t T237_2	p T280_2	
b T268_2	p T279_1	p T237_3	t T281	
b T269_1	p T279_2	d[b] T243		
b T271		j T249_1		
b T274		j T249_2		
b T275		t T251		
b T276		j T262_2		
b T277		i T263		
p T280_1				
b T282				

[†]Targets are separated into columns based on the classification of the assessors. The method used by Robetta to model the domain is indicated next to the target, with "b": parent confidently detected by BLAST, "p": parent confidently detected by PSI BLAST, "f": parent confidently detected by FFAS03, "i": intermediate FFAS03 hit (used 3D-Jury for parent), "n": intermediate FFAS03 hit, but low confidence 3D-Jury (modeled de novo), "j": parent confidently detected by 3D-Jury, "t": parent detected by 3D-Jury, but with a "twilight-zone" confidence (modeled both template-based and de novo), "d": modeled de novo.

[a]Robetta_04 detected a new PSI-BLAST hit that became available between the release of the target sequence and the expiration of the target

[b]Robetta_04 used new twilight-zone 3D-Jury hit and also modeled de novo.

TABLE II. Average GDT_TS for each Method categorized by Detection Regime[†]

		Base		Rob		Rob_04		Baker	
Regime	N	First	Best	First	Best	First	Best	First	Best
De novo	14	N/A	N/A	30.7	33.2	29.0	34.3	39.1	41.4
3D-Jury	23	31.3	34.0	34.3	41.2	36.8	41.8	42.6	46.7
FFAS03	6	47.8	49.4	46.8	51.7	50.6	52.4	54.4[a]	59.0[a]
PSI-BLAST	44	62.1	62.5	63.3	66.8	62.8	65.8	65.1	67.2

[†]The Average GDT_TS score for the first and best model from the "Base" detection method (3D-Jury, FFAS03, or BLAST/PSI-BLAST), the Robetta server, the Robetta_04 protocol, and the Baker human group. The number of domain targets evaluated in each category is shown. Only those models that were generated using parents from the same detection method were used in the calculation, leading to the exclusion from the Baker human Best calculation for the de novo regime of the template-based model 5 for target T0213, T0214, and T0227, the exclusion from the Best calculations for the de novo regime of template-based models for target T0243 of Baker human model 5 and Robetta_04 models 1, 3, and 5, the exclusion of T0216 domains 1 and 2 from the de novo regime altogether as the detection level changed between the execution time of Robetta and Robetta_04, the exclusion of target T0226 domain 1 from the 3D-Jury regime altogether as the detection level changed.
[a]The Baker human group did not submit a model for target T0228 domains 1 and 2, and therefore the average for the FFAS03 regime for the Baker human group is with respect to four targets instead of six.

improvement provided by the application of our methods over the baseline method is for the more remote targets. This reflects the challenge in improving the quality of the already quite close models derived from evolutionarily close parents. Both of the automated methods and the human-intervention method show improved modeling over the baseline detection method in the FFAS03 and 3D-Jury Fold Recognition regime. We also find that the Baker human group does a much better job predicting native-like conformations in the de novo regime than the automated methods.

What Went Right

The Robetta and Robetta_04 methods were successful in producing models that were better, on average, than those from the baseline detection methods in the Fold Recognition regime, often approaching, and sometimes exceeding, the quality of our human group's models. Additionally, while the automatic methods were not as successful as our human group's ability to select the superior model to submit as the first model, the best models produced by the automatic de novo protocol were often comparable to the human group's best de novo models.

The results for each target in the Fold Recognition regime are shown in Figure 1. Figure 1(a) shows the difference in GDT_TS between the model from the detection server that was the top confidence detection (or longest detection of similar confidence) and the first model from the Robetta server, the Robetta_04 protocol, and our human group. Figure 1(b) shows this same information, but compares the best model from each of our methods with the best model from the detection server (only considering those parents that were utilized by our methods). Figure 1(c) shows the absolute GDT_TS score achieved by the first and best detection model for each target. We see that our methods often either maintain or improve over the baseline detection models. The more challenging targets tend to be those that our modeling approaches most improve, whereas those targets that are already quite well predicted by the detection method sometimes suffer from continued modeling.

Some of the automatic template-based modeling highlights are shown in Figure 2. The Robetta server provided among the best first models for all groups for targets T0206 and T0251 (shown), whereas Robetta_04 offered among the best first models for targets T0262 domain 2, T0248 domain 1, T0262 domain 2, T0263 (shown), and T0281 (shown). The improvement in quality for many of the targets provided by the Robetta_04 method over the Robetta server models implies that the use of multiple parents and backbone refinement for template-based modeling in the Fold Recognition regime often provides worthwhile contributions, and we will therefore incorporate these concepts into the server.

Automatic de novo modeling using the Robetta protocol was also competitive with methods used by many human predictors in CASP6. Figure 3(a) shows the first model GDT_TS scores for each of the targets that were initially classified by Ginzu as purely de novo targets for Robetta, Robetta_04, and the Baker (Human) group. Figure 3(b) indicates the GDT_TS scores of the best of the five submitted models. Some of the automatic de novo highlights are shown in Figure 4. Highlights included Robetta first models for targets T0209 domains 1 and 2 (shown), and alternate models for targets T0230 (model 2 shown, which as a twilight-zone target was the first de novo model), T0248 domain 1 (model 2 shown, the first de novo model), and T0248 domain 2 (model 2, the first de novo model). Robetta_04 used the same cluster centers as Robetta from the de novo decoy ensemble, but reranked them using a confidence score[18] that includes MAMMOTH[45] matches to known structures. Robetta_04 de novo highlights included first models for T0272 domain 2 and T0273, and alternate models for targets T0209 domain 1 (model 3), T0230 (model 2, the first de novo model), T0241 domain 1 (model 2, the first de novo model), T0248 domain 1 (model 2, the first de novo model), and T0272 domain 1 (model 5). While automated de novo modeling was not quite as good as modeling employing human intuition in either the Fold Recognition/Analogous or the New Fold categories, reasonable models were often produced among the five. It was our hope that other groups would feel free to use Robetta models as starting points for continued modeling, or perhaps apply a discrimination method to select the best model from the five better than our own efforts, and we were delighted to learn at the

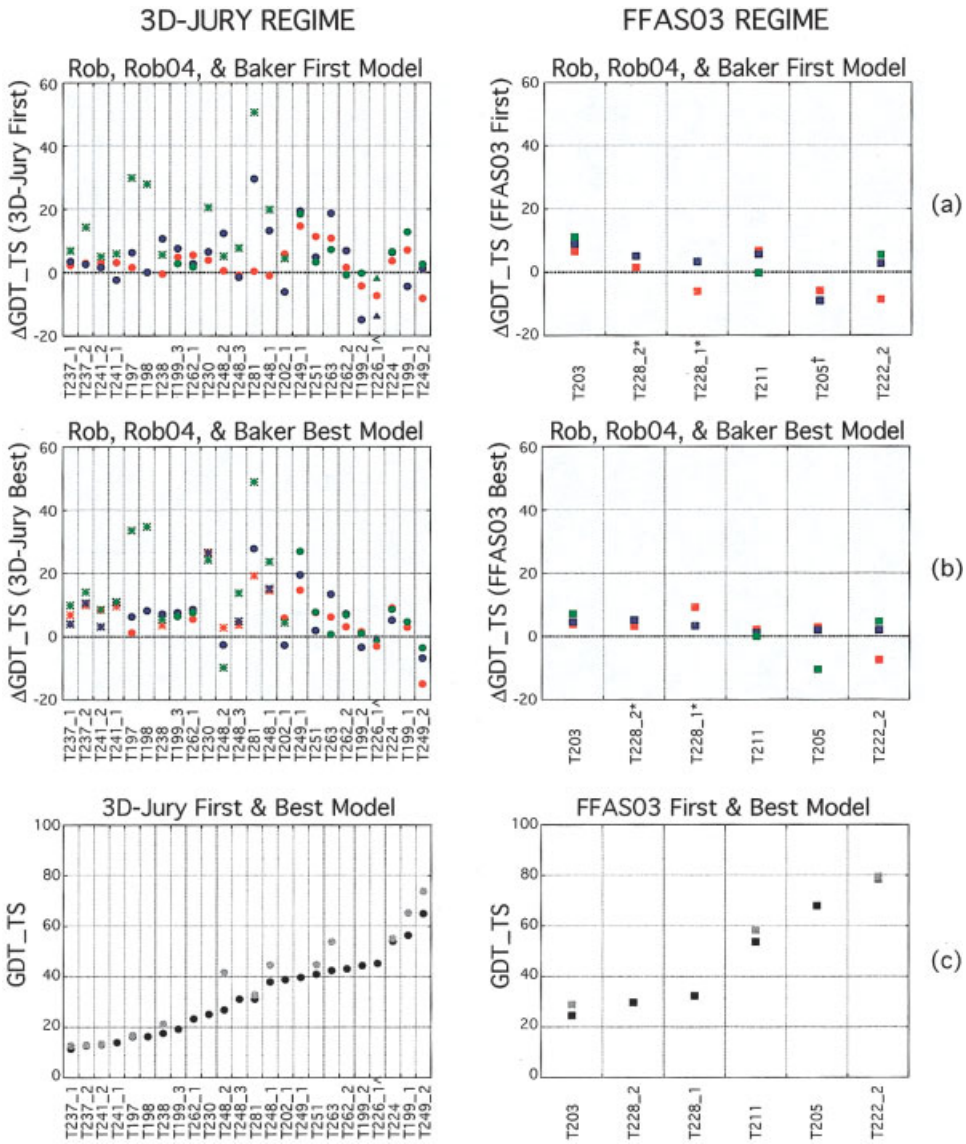

Fig. 1. Target-specific model quality for Robetta, Robetta_04, and Baker Human Group in the 3D-Jury and FFAS03 regimes. **a**: The difference in GDT_TS score between the First model of the detection method (either 3D-Jury or FFAS03) First model and the First model of Robetta ("Rob", in red), Robetta_04 ("Rob_04", in blue) and the Baker human group ("Baker", in green). De novo models are indicated with an asterisk, template-based models built from parents detected by 3D-Jury shown with circles, template-based models from FFAS03 detections with squares, and template-based models from PSI-BLAST detections with triangles. **b**: The difference in GDT_TS score between the detection method Best model (only considering those parents that were used for modeling by our methods) and the Robetta, Robetta_04, and Baker human group Best models. Colors and point types are as before. **c**: Absolute GDT_TS score of the detection method First model (in black) and Best model (in gray, shown only if different from the First model). *: The Baker human group did not submit models for target T0228. ^: Target T0226 changed from a 3D-Jury detection to a PSI-BLAST detection (which did not appear to provide a superior parent!) in the time intervening the release of the target sequence and the due date of the models. †: The Baker human group first model for target T0205 was too low-scoring to display on the plot, receiving a ΔGDT_TS score of − 23.

CASP6 meeting that the models had indeed been useful to other predictors.

What Went Wrong

All modeling when performed for domains separately is dependent on the accuracy with which the domain boundaries are assigned. Robetta was hindered by inaccuracies in the Ginzu parse for targets T0208 and T0235, where the PSI-BLAST alignments caused an incomplete coverage of the query, leading to an excessive number of domains. We expect to remedy the problem encountered by target T0235 by adding logic to recognize when the putative domains of the protein belong to substructures of folds in the same SCOP[46] class.

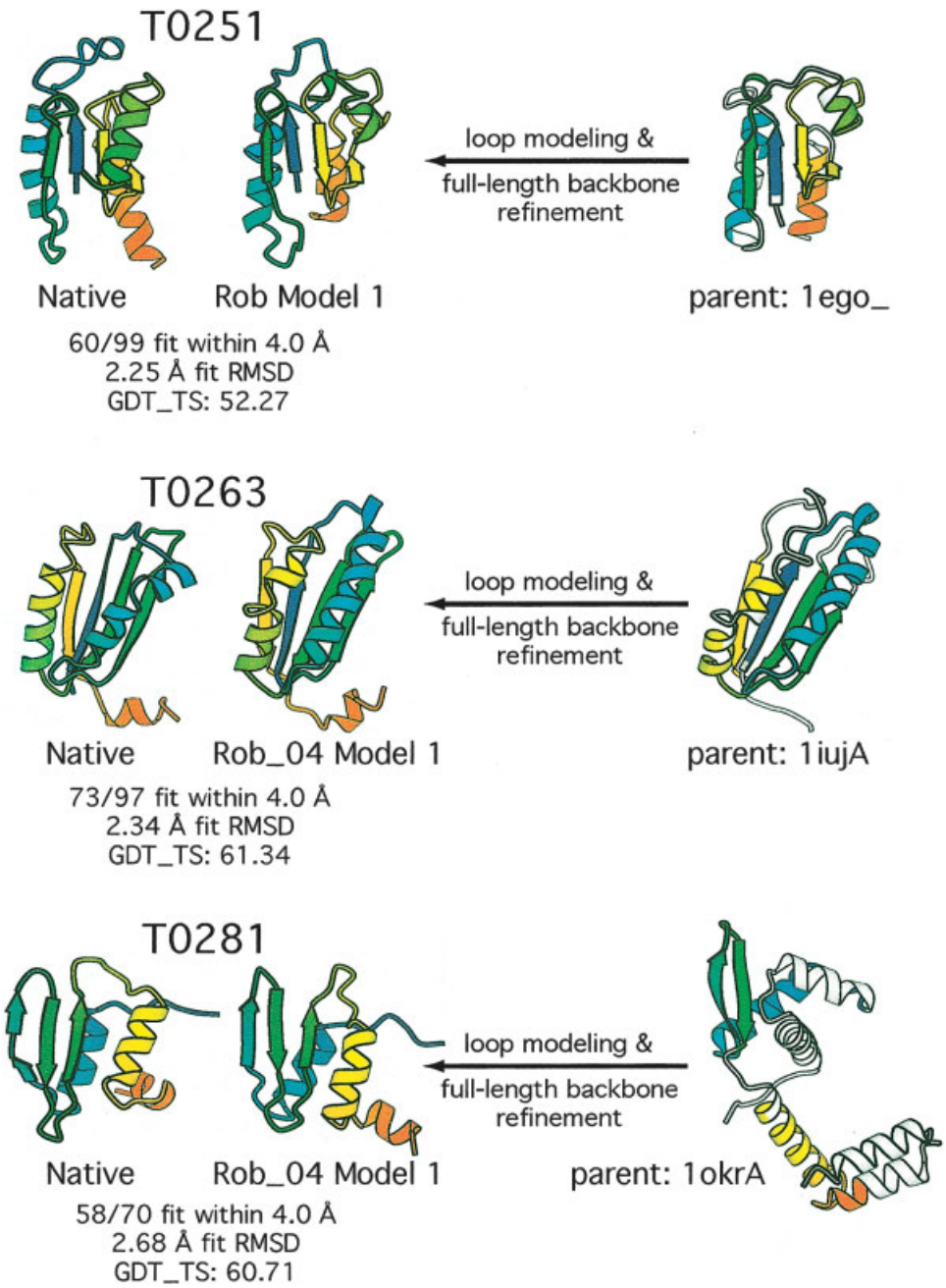

Fig. 2. Robetta and Robetta_04 template-based model highlights. The first model from Robetta ("Rob") for target T0251 and the first models from Robetta_04 ("Rob_04") for targets T0263 and T0281 are shown in cartoon representation. The native and the model are depicted with a rainbow coloring, with blue at the N-terminus and red at the C-terminus. The structure of the parent is shown with unaligned positions in white and aligned regions that produced the initial template colored to indicate the corresponding positions in the final model. The fraction of residues in the native structure that deviate no more than 4.0 Å from the model in a sequence-dependent optimal superposition by LGA, the RMSD of the LGA fit, and the GDT_TS score of the model are reported. Note the missing density between the green strand and the green helix in target T0251's native structure which has been also removed from the model for clarity. Also note the adjustment to the model of the yellow and turquoise helices from 1iujA in target T0263 and the dramatic shift in the model for target T0281 of the yellow helix from 1okrA to form a more compact structure.

Target T0240 possessed a strong BLAST hit, and we therefore utilized that parent for modeling with both Robetta and Robetta_04. However, human examination of putative parents detectable from a PSI-BLAST search revealed a more distant relative that suggests the possibility that the β-hairpin swap found in the BLAST homodimer is not obligatory, and indeed the β-hairpin swap does not occur in the native structure of T0240. Our

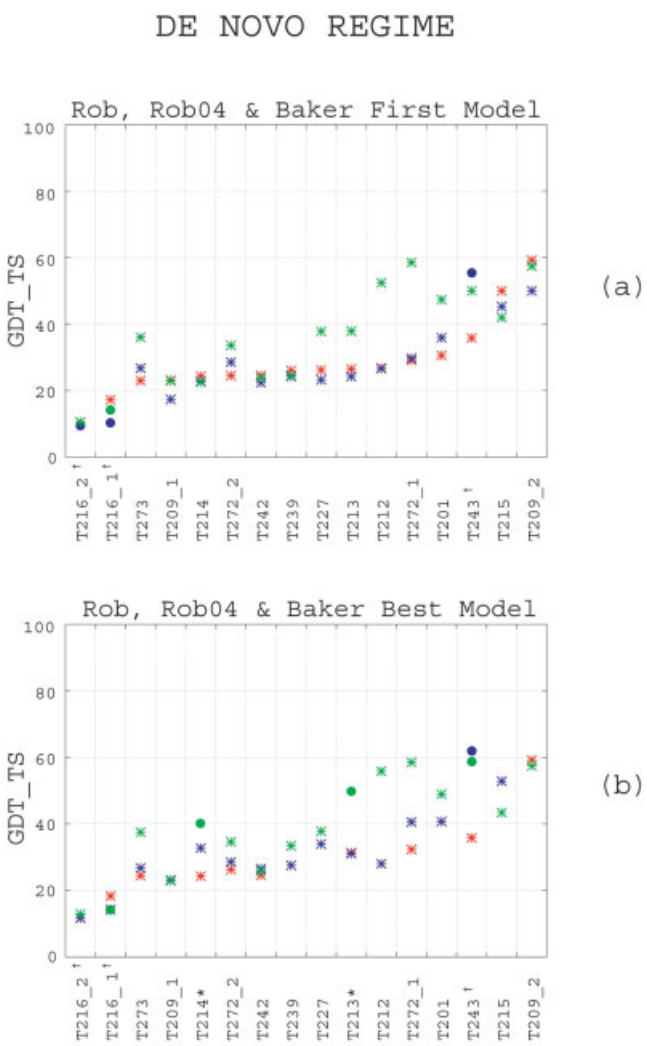

Fig. 3. Target-specific model quality for Robetta, Robetta_04, and Baker human group in the de novo regime. **a**: The absolute GDT_TS score of the First model of Robetta (in red), Robetta_04 (in blue), and the Baker human group (in green). De novo models are indicated with an asterisk, and template-based models with a circle. **b**: The absolute GDT_TS score of the Best model of Robetta, Robetta_04, and the Baker human group, with colors and point types as before. *: the best Baker human models for targets T0213 and T0214 were template-based models using a parent suggested by Alexey Murzin. †: the 3D-Jury confidence changed between the release of the target sequence and the due date of each target for targets T0216 and T0243, leading these targets to be considered "twilight-zone" targets and modeled by both template-based and de novo protocols by Robetta_04, and using a template-based approach by the Baker human group.

automated methods were unable to account for this conformational flexibility in the context of the dimer as all modeling is performed with the assumption that the target is a monomer.

Unfortunately, we suffered a hardware failure shortly after the completion of the CASP6 prediction period, leading to a total loss of all data from the intermediate stages of the Robetta_04 template-based modeling and leaving us with only the final submitted models. This has impaired our ability to distinguish precisely the factors responsible for the differences between the Robetta and Robetta_04 template-based modeling protocols. However, based on recent work with another benchmark (D.C., manuscript in preparation), the gain in performance on the Fold Recognition targets likely results primarily from the inclusion of parametric alignment ensembles from multiple parents.

During CASP6, selection of models from alignment ensembles was performed purely based on energy. Recent findings (D.C., manuscript in preparation) have shown selection to be greatly enhanced by additional evaluation of alignments from the ensemble using a consensus approach. We expect this would have had the greatest impact on Robetta's model quality in the Fold Recognition regime, hopefully making them even better.

Several Fold Recognition targets had a suitable parent that was recommended by Alexey Murzin on the FORCASP website (http://www.forcasp.org/), specifically targets T0213, T0214, and T0227. The Ginzu logic found an intermediate FFAS03 parent for targets T0213 and T0214, but it was not supported by the 3D-Jury results at the time of execution. Target T0227 did not have a confident parent detected by either FFAS03 or 3D-Jury. All three targets were therefore de novo modeled by Robetta and Robetta_04, leading to a poor relative performance on these targets as compared with groups that either found the fold by other means or took the hint, illustrating the difficulty one has in comparing automated methods with groups which make use of human intuition and expertise.

New Fold targets were frequently modeled with accuracy comparable to other methods, sometimes possessing the correct topology, but failed to capture the high-resolution features that are ultimately the goal of protein structure prediction.Additionally, as illustrated by Robetta's failure on target T0215, fragment-based methods are dependent on the quality of the fragments in the library. In this case, the fragment library used by Robetta possessed large quantities of strand conformations near the C-terminus, causing the top decoy clusters to contain models with a C-terminal hairpin that should have been a helix.

The effort by Robetta_04 to improve selection of the first de novo model was not a huge success, with perhaps only targets T0214 and T0272 exhibiting any tendency towards improved selection. Since the confidence function is dominated by the clustering threshold (which is the same for both methods) and by any MAMMOTH detections, it should not be surprising that the successes are found for Fold Recognition targets that possess topologically similar experimental structures, with less difference for the true New Fold targets.

What We Learned

It was challenging to improve on models for the easiest Comparative Modeling targets. However, Robetta was more successful with the more remote Fold Recognition and New Fold targets. We found it possible to improve our automatic template-based modeling protocol, especially in the Fold Recognition regime, where the greatest room for improvement exists in the identification of the best parent and the alignment to it, as well as refinement of the

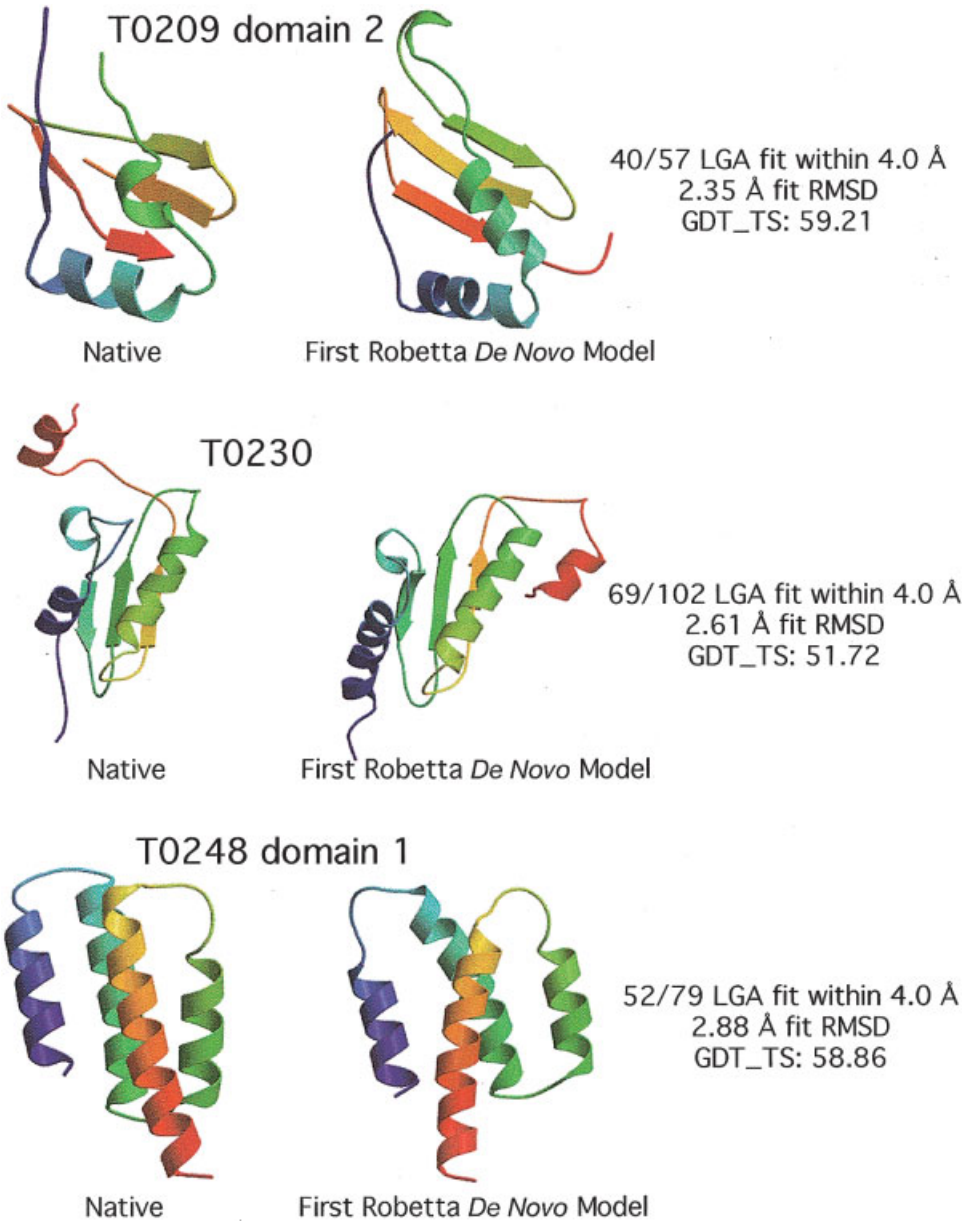

Fig. 4. Robetta de novo model highlights. The first Robetta de novo model and the native structure are shown in a cartoon representation for targets T0209 domain 2, T0230, and T0248 domain 1, where blue indicates the N-terminus and red the C-terminus. The fraction of residues found in the native that is fit within 4 Å by the model using a sequence-dependent LGA structural superposition is reported, as is the RMSD of the fit and the GDT_TS score achieved by the model. The model for target T0209 domain 2 retains the residues that were missing density in the native structure to clarify the topology. Targets T0230 and T0248 were classified by Robetta as twilight-zone 3D-Jury targets, so were modeled with both template-based and de novo protocols. The de novo models for targets T0230 and T0248 domain 1 were much better than the template-based models from Robetta.

backbone. Less successful, but clearly possible based on the superior results by our human group for some of the most challenging targets, were efforts to find a better protocol for the automatic de novo modeling of targets. Early results have shown some promise for pseudo-de novo modeling of twilight-zone targets by employing predicted distance restraints or long-range strand pairing from low-confidence models reported by Fold Recognition servers (see accompanying Baker human group report in this issue), and we will investigate the feasibility of incorporation of such methods into the server for the modeling of the most remote Fold Recognition targets.

In the near term, the models provided by automated modeling remain coarse approximations of the native structure, but may be used as starting points for further modeling. For example, as discussed in our human group's report in this issue, high-resolution refinement protocols may potentially be applied to template-based and de novo

models that possess the correct topology in order to obtain more native-like structures. Provided that the computing power becomes available, we hope to enable high-resolution refinement in the Robetta server, with the ultimate goal of providing fully automatic high-resolution modeling for any and all protein sequences.

ACKNOWLEDGMENTS

The authors thank the structural biologists for allowing their structures to be used in CASP, the CASP organizers and assessors for implementing the CASP6 experiment, and Dani Fischer for running the CAFASP experiment. We thank Kevin Karplus for the use of the SAM-T99 software, David Jones for the use of the PSIPRED software, Jens Meiler for the use of the JUFO software, Adam Godzik for the use of the FFAS03 server, William Taylor for the use of his structure-based domain identification program, and Leszek Rychlewski for the use of the 3D-Jury BioInfo meta server and the 3D-pair program. We also specially thank all server developers, whose collective work is so essential to consensus and consensus-derivative methods such as Robetta. We strongly feel that, like the CASP experiment itself, the success of such methods shows that the science of protein structure prediction can best be furthered by our working together as a community. The authors also thank Keith Laidig for effective and innovative administration and design of the Robetta hardware resources, as well as Richard Bonneau and Charlie Strauss for their generous provision of the Robetta hardware. D.C. and D.K. are supported by the NIH/NIGMS SGPP initiative. This work was also supported by the HHMI.

REFERENCES

1. Kinch LN, Wrabl JO, Krishna SS, Majumdar I, Sadreyev RI, Qi Y, Pei J, Cheng H, Grishin NV. CASP5 assessment of fold recognition target predictions. Proteins 2003;Suppl 6:395–409.
2. Aloy P, Stark A, Hadley C, Russell RB. Predictions without templates: new folds, secondary structure, and contacts in CASP5. Proteins 2003;Suppl 6:436–456.
3. Fischer D, Rychlewski L, Dunbrack Jr. RL, Ortiz AR, Elofsson A. CAFASP3: the third critical assessment of fully automated structure prediction methods. Proteins 2003;Suppl 6:503–516.
4. Chivian D, Kim DE, Malmstrom L, Bradley P, Robertson T, Murphy P, Strauss CE, Bonneau R, Rohl CA, Baker D. Automated prediction of CASP5 structures using the Robetta server. Proteins 2003;Suppl 6:524–533.
5. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 2004;32(Web Server issue):W526–531.
6. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 1997;268:209–225.
7. Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins 1999;34:82–95.
8. Goldberg ME. Tertiary structure of Escherichia coli beta-D-galactosidase. J Mol Biol 1969;46:441–446.
9. Wetlaufer DB. Nucleation, rapid folding, and globular intrachain regions in proteins. Proc Natl Acad Sci USA 1973;70:697–701.
10. Richardson JS. The anatomy and taxonomy of protein structure. Adv Protein Chem 1981;34:167–339.
11. 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.
12. Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res 2001;29:2994–3005.
13. Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A. FFAS03: a server for profile—profile sequence alignments. Nucleic Acids Res 2005;33(Web Server issue):W284-8.
14. Rychlewski L, Jaroszewski L, Li W, Godzik A. Comparison of sequence profiles. Strategies for structural predictions using sequence information. Protein Sci 2000;9:232–241.
15. Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 2003;19:1015–1018.
16. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, et al. The Pfam protein families database. Nucleic Acids Res 2004;32(Database issue):D138–141.
17. Bonneau R, Tsai J, Ruczinski I, Chivian D, Rohl C, Strauss CE, Baker D. Rosetta in CASP4: progress in ab initio protein structure prediction. Proteins 2001;Suppl 5:119–126.
18. Bonneau R, Strauss CE, Rohl CA, Chivian D, Bradley P, Malmstrom L, Robertson T, Baker D. De novo prediction of three-dimensional structures for major protein families. J Mol Biol 2002;322:65–78.
19. Carugo O, Pongor S. A normalized root-mean-square distance for comparing protein three-dimensional structures. Protein Sci 2001;10:1470–1473.
20. Bradley P, Chivian D, Meiler J, Misura KM, Rohl CA, Schief WR, Wedemeyer WJ, Schueler-Furman O, Murphy P, Schonbrun J, et al. Rosetta predictions in CASP5: successes, failures, and prospects for complete automation. Proteins 2003;Suppl 6:457–468.
21. Waterman MS. Parametric and ensemble sequence alignment algorithms. Bull Math Biol 1994;56:743–767.
22. Jaroszewski L, Li W, Godzik A. In search for more accurate alignments in the twilight zone. Protein Sci 2002;11:1702–1713.
23. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 1999;292:195–202.
24. Karplus K, Barrett C, Hughey R. Hidden Markov models for detecting remote protein homologies. Bioinformatics 1998;14:846–856.
25. Meiler J, Mueller M, Zeidler A, Schmaeschke F. JUFO: secondary structure prediction for proteins. http://www.jens-meiler.de/; 2002.
26. Rohl CA, Strauss CE, Chivian D, Baker D. Modeling structurally variable regions in homologous proteins with rosetta. Proteins 2004;55:656–677.
27. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using Rosetta. Methods Enzymol 2004;383:66–93.
28. John B, Sali A. Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res 2003;31:3982–3992.
29. Contreras-Moreira B, Fitzjohn PW, Bates PA. In silico protein recombination: enhancing template and sequence alignment selection for comparative protein modelling. J Mol Biol 2003;328:593–608.
30. Petrey D, Xiang Z, Tang CL, Xie L, Gimpelev M, Mitros T, Soto CS, Goldsmith-Fischman S, Kernytsky A, et al. Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. Proteins 2003;Suppl 6:430–435.
31. Dunbrack RL Jr., Cohen FE. Bayesian statistical analysis of protein side-chain rotamer preferences. Protein Sci 1997;6:1661–1681.
32. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci USA 2000;97:10383–10388.
33. Canutescu AA, Dunbrack RL Jr., Cyclic coordinate descent: a robotics algorithm for protein loop closure. Protein Sci 2003;12:963–972.
34. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 1992;89:10915–10919.
35. 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.
36. McGuffin LJ, Jones DT. Improvement of the GenTHREADER

method for genomic fold recognition. Bioinformatics 2003;19:874–881.
37. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 1999;287:797–815.
38. Shi J, Blundell TL, Mizuguchi K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol 2001;310:243–257.
39. Kelley LA, MacCallum RM, Sternberg MJ. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 2000;299:499–520.
40. Fischer D. Hybrid fold recognition: combining sequence derived properties with evolutionary information. Pac Symp Biocomput 2000:119–130.
41. Ginalski K, von Grotthuss M, Grishin NV, Rychlewski L. Detecting distant homology with Meta-BASIC. Nucleic Acids Res 2004;32(Web Server issue):W576–581.
42. Taylor WR. Protein structural domain identification. Protein Eng 1999;12:203–216.
43. Plewczynski D, Pas J, von Grotthuss M, Rychlewski L. 3D-Hit: fast structural comparison of proteins. Appl Bioinformatics 2002;1:223–225.
44. Zemla A. LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res 2003;31:3370–3374.
45. Ortiz AR, Strauss CE, Olmea O. MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. Protein Sci 2002;11:2606–2621.
46. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.