Protein Structure Prediction by Pro-Sp3-TASSER

Hongyi Zhou and Jeffrey Skolnick*
Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology, Atlanta, Georgia

ABSTRACT An automated protein structure prediction algorithm, pro-sp3-Threading/ASSEmbly/Refinement (TASSER), is described and benchmarked. Structural templates are identified using five different scoring functions derived from the previously developed threading methods PROSPECTOR_3 and SP^3. Top templates identified by each scoring function are combined to derive contact and distant restraints for subsequent model refinement by short TASSER simulations. For Medium/Hard targets (those with moderate to poor quality templates and/or alignments), alternative template alignments are also generated by parametric alignment and the top models selected by TASSER-QA are included in the contact and distance restraint derivation. Then, multiple short TASSER simulations are used to generate an ensemble of full-length models. Subsequently, the top models are selected from the ensemble by TASSER-QA and used to derive TASSER contacts and distant restraints for another round of full TASSER refinement. The final models are selected from both rounds of TASSER simulations by TASSER-QA. We compare pro-sp3-TASSER with our previously developed MetaTASSER method (enhanced with chunk-TASSER for Medium/Hard targets) on a representative test data set of 723 proteins $<$250 residues in length. For the 348 proteins classified as easy targets (those templates with good alignments and global structure similarity to the target), the cumulative TM-score of the best of top five models by pro-sp3-TASSER shows a 2.1% improvement over MetaTASSER. For the 155/220 medium/hard targets, the improvements in TM-score are 2.8% and 2.2%, respectively. All improvements are statistically significant. More importantly, the number of foldable targets (those having models whose TM-score to native $>$0.4 in the top five clusters) increases from 472 to 497 for all targets, and the relative increases for medium and hard targets are 10% and 15%, respectively. A server that implements the above algorithm is available at http://cssb.biology.gatech.edu/skolnick/webservice/pro-sp3-TASSER/. The source code is also available upon request.

INTRODUCTION

Although the past few years have seen slow and continuous progress, the protein structure prediction problem remains a long-standing unsolved problem in structural biology (1,2). In general, protein structure prediction approaches can be classified into template-based and template-free. To date, template-based approaches are still the only reliable method for protein structure prediction (1,2). Template-based methods generally involve template identification, alignment of the target sequence to the template, and model refinement from this initial alignment. Advances can be seen in better template identification and sequence-to-structure alignment by going from single sequence alignments to profile alignments (3–11), by inclusion of multiple structural properties such as predicted secondary structures (12–14), solvent accessibility (15,16), and structure profiles (17–24); by coupling alignment with quality assessment (25); machine learning (26,27), and meta-servers (28–30). Better model refinement can be achieved by using multiple templates and multiple template alignments (31–33), iterative refinement (34–36), as well as physics and evolution-based potentials (33,37,38).

Although not as reliable as template-based approaches, template-free or ab initio approaches have had significant advances (36,37,39–45). These methods can occasionally provide very good models for targets where no related proteins in the template library can be identified. There are studies which show that the current solved structural library Protein Data Bank (PDB) (46) could provide reasonable quality templates for all compact single domain proteins (47–49). Despite the progress made by many efforts, there are still ~1/3 of proteins that are weakly homologous to proteins in the PDB for which appropriate templates cannot be identified (1,2). Therefore, in practice, there are two ways to solve the protein structure prediction problem: improving the reliability of ab initio predictions and/or the sensitivity of template identification. The protein structure prediction method TASSER (Threading/ASSEmbly/Refinement) (50) and its updated version chunk-TASSER (44) for medium/hard targets have achieved a reasonable level of success in both template-based and template-free modeling for targets that are weakly homologous to template proteins (50,51). Their performance depends strongly on the energy functions that are partly derived from the input structures. The input structures can be models built directly from target-template alignments, or models generated by other methods, and ab initio folded chunk structures in the case of chunk-TASSER. One way to improve the energy functions is to improve the predicted contact potentials as in iterative TASSER (35) and TASSER 2.0(52). The other way to improve the accuracy of TASSER model is to use more accurate input structures as in MetaTASSER (51) which uses the 3D-jury algorithm (28) to select input models from three state-of-the-art threading methods.

Submitted September 12, 2008, and accepted for publication December 3, 2008.

*Correspondence: skolnick@gatech.edu

Editor: Marcia Newcomer.

© 2009 by the Biophysical Society

0006-3495/09/03/2119/9 $2.00

doi: 10.1016/j.bpj.2008.12.3898

In this study, our goal is to improve protein structure prediction accuracy by improving the accuracy of the models that are inputted into TASSER or chunk-TASSER. The accuracy of the input models depends on the quality of the identified templates and the associated target-template alignments as generated by threading methods. In this spirit, we have developed what to our knowledge is a new in-house meta-threading method that consists of five individual threading algorithms and alternative alignment procedures that employ a parametric alignment approach. We then use short TASSER or chunk-TASSER to build models from threading templates and subsequently select possibly better models by a model quality assessment method. This procedure can be considered as iterative metathreading with the first round selecting models from five individual threading methods and the second round selecting models from short TASSER or chunk-TASSER predicted models. The selected models are then refined by full TASSER or chunk-TASSER simulations. On a benchmark set of 723 targets, we show that the method developed here shows considerable improvement over our previously developed MetaTASSER algorithm (51) in terms of the number of foldable targets and on average a 2–3% improvement in TM-score (53).

METHODS

Employing the TASSER and chunk-TASSER (33,44,50,51), the model quality assessment method TASSER-QA (54), and the threading methods, PROSPECTOR_3 (55) and SP^3(24), we have developed the pro-sp3-TASSER algorithm that combines some of the above-mentioned successful approaches such as meta-servers, coupling alignment with quality assessment, and iterative refinement. The overall pro-sp3-TASSER procedure is described by the flowchart shown in Fig. 1 *A* and has three main steps: a), threading and alignment by PRO-SP^3 (described below); b), generation of an ensemble of models by short TASSER simulations; and c), selection of models by TASSER-QA, followed by full TASSER refinement and final model selection. In addition to the SP^3 threading method, we derived four other threading methods by modifying and combining some of the score components in SP^3 and PROSPECTOR_3. Combination of these threading methods provides consensus information from individual threading/fold-recognition approaches to improve the quality of template identification and alignment accuracy, especially for medium/hard targets (those with poor quality template identification/alignments). Additional alternative target sequence to template alignments are generated by a parametric alignment method where top quality alignments are selected by TASSER-QA for medium/hard targets. Subsequently, limited short time TASSER refinement is applied to generate an ensemble of full-length models. For medium/hard targets, chunk-TASSER (44) refinement, which uses ab initio folded supersecondary structure chunks of a given target, was also used instead of the original TASSER algorithm. The structure ensemble consists of up to 155 models and provides a diverse quality of models. We then select top 20 models from the ensemble using TASSER-QA. A second round of full TASSER refinement was applied on the selected full-length models and final models were selected from both rounds of TASSER or chunk-TASSER refinements.

Target difficulty is classified by the *Z*-score of the top template in SP^3 threading (part of PRO-SP^3). Targets whose top template has a *Z*-score ≥ 6.0 are classified as easy, those with a *Z*-score $\leq$ 4.5 as hard, and those having a $4.5 <$ *Z*-score < 6.0 as medium targets, respectively. Special attention is paid to possible multiple domain targets. This is done by first checking the coverage of the top template identified by SP^3. If more than 50 continuous residues are unaligned in the top scoring template, in additional to modeling the full-length target sequence, the unaligned and aligned regions are modeled separately. The separately modeled possible domains are then superimposed onto the full-length models in the second round of TASSER refinement. Another special case is that when a single template ranked top by SP^3 score has a *Z*-score more than 2.0 units larger than the rank two template or if it has sequence identity to the target of more than 50% (single template dominates), then we only use the single, top scoring template in the subsequent modeling procedure.

PRO-SP^3 threading for template identification and sequence-to-structure alignment

The PRO-SP^3 threading algorithm consists of five threading scores that align the target sequence to templates and then ranks the templates independently. These scores are taken directly or modified from SP^3 (24) and PROSPECTOR_3 (55). The first score is taken directly from SP^3 (24) and combines sequence profiles from sequence and structural alignments:

$$\begin{aligned} S_1(i,j) \;=\; & -(1-w_{\text{struc}})F^{\text{seq}}_{\text{query}}(i)\cdot M^{\text{seq}}_{\text{template}}(j) \\ & - w_{\text{struc}}F^{\text{struc}}_{\text{template}}(j)\cdot M^{\text{seq}}_{\text{query}}(i) - w_{\text{2ndary}}\delta_{s_i,s_j} \\ & + S_{\text{shift}}, \end{aligned} \tag{1}$$

where $F^{\text{seq}}_{\text{query}}(i)$ is the sequence-derived frequency profile of the target sequence, $M^{\text{seq}}_{\text{template}}(j)$ is the sequence-based log odds profile (position-specific substitution matrix as in PSIPRED (56)) of the template, $F^{\text{struc}}_{\text{template}}(j)$ is the structure-derived probability profile of the template (24), $M^{\text{seq}}_{\text{query}}(i)$ is the sequence-derived, log odds profile of the target sequence, S_{shift} is a constant shift factor, w_{struc} and w_{2ndary} are weight parameters for the structure-derived sequence profiles and secondary structure profiles respectively, and δ_{s_i,s_j} is a simple function of the secondary structure element s_i of the target at sequence position i and s_j of the template at sequence position j:

$$\delta_{s_i,s_j} = \begin{cases} 1 & si = sj \\ -1 & si \neq sj \end{cases}. \tag{2}$$

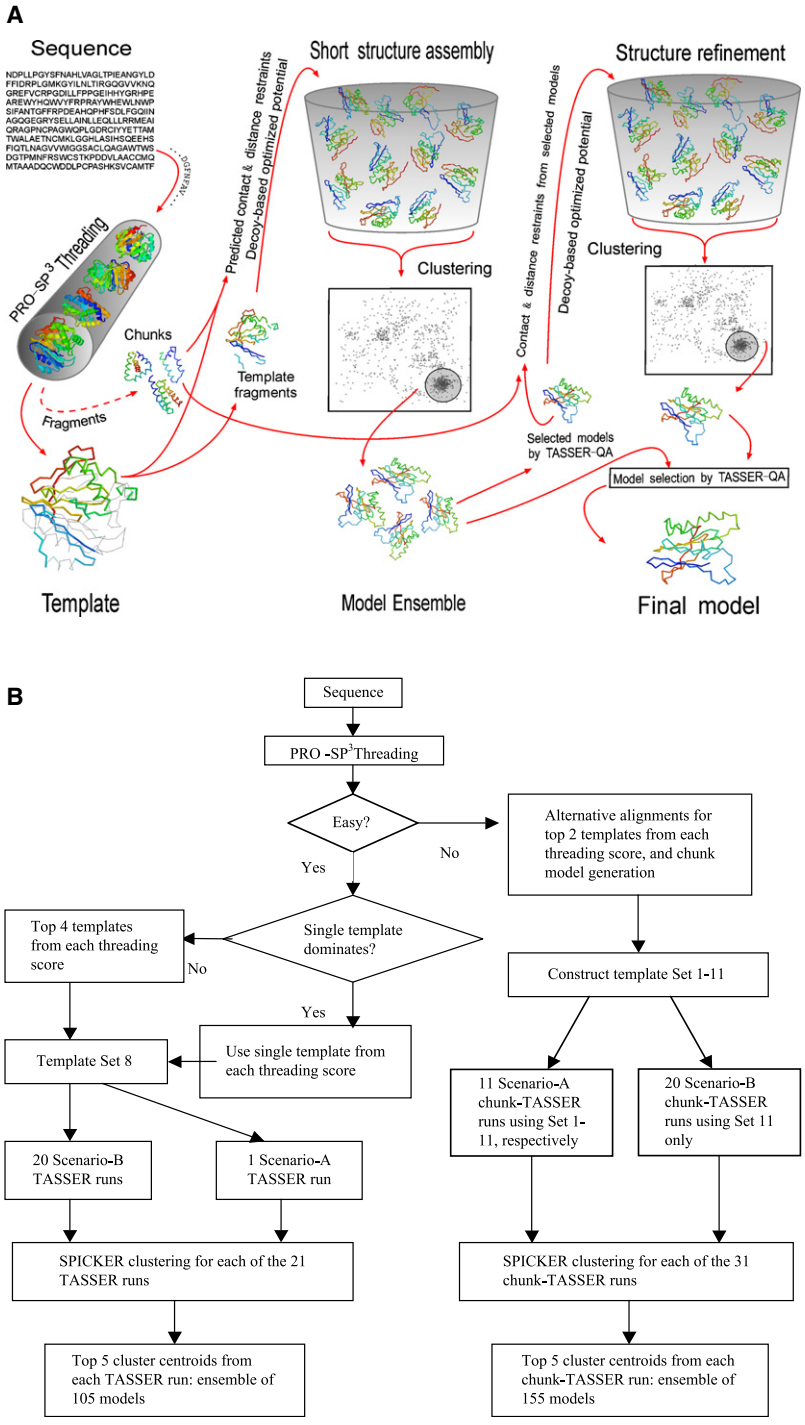

FIGURE 1 (*A*) Flowchart of pro-sp3-TASSER. (*B*) Detailed flowchart of model ensemble generation.

Notice that the secondary structure of the target sequence is predicted by PSIPRED (56) using the same position-specific substitution matrix (PSSM), $M^{\text{seq}}_{\text{query}}(i)$. In the latest version of SP^3, $F^{\text{seq}}_{\text{query}}(i)$ is replaced by the average of two profiles with PSIBLAST e-value cutoffs 0.001 and 1.0, and $F^{\text{struc}}_{\text{template}}(j)$ is the average of the original $F^{\text{struc,org}}_{\text{template}}(j)$ and the PSIBLAST profile with e-value cutoff 0.001; i.e., $F^{\text{seq}}_{\text{query}}(i) = (F^{\text{seq,0.001}}_{\text{query}}(i) + F^{\text{seq,1.0}}_{\text{query}}(i))/2$, and $F^{\text{struc}}_{\text{template}}(j) = (F^{\text{struc,org}}_{\text{template}}(j) + F^{\text{seq,0.001}}_{\text{template}}(j))/2$.

These modifications are designed to increase the alignment accuracy of medium/hard targets. With the gap opening and extension parameters (w_0, w_1), there are total of five parameters which were determined by optimizing alignment accuracy against the ProSup data set (57). The parameters were reoptimized after the profile replacements mentioned above and the new solution (w_0, w_1, $w_{2\text{ndary}}$, w_{struc}, S_{shift}) is (3.5, 0.1, −1.50, 0.5, 0.7). The new optimized set of weights gave a one-to-one match alignment accuracy of 66.1%

against the ProSup structure alignment benchmark (57) compared to the original accuracy of 65.3%.

Scores 2–5 are modified from PROSPECTOR_3 by incorporating the structure-derived profiles from SP^3 into PROSPECTOR_3. This is done because structure-derived profiles help increase the template identification sensitivity (16,24). The scoring functions given below are different combinations of sequence profiles and structure-derived profiles beyond the SP^3 approach. At the same time, they take into account the well characterized PROSPECTOR_3 (55) pair potentials.

$$S_2(i,j) = -\big(F^{\mathrm{seq}}_{\mathrm{query}}(i)\cdot M^{\mathrm{str,org}}_{\mathrm{template}}(j) + F^{\mathrm{struc}}_{\mathrm{template}}(j)\cdot M^{\mathrm{seq}}_{\mathrm{query}}(i)\big)/2 - W^2_{\mathrm{2ndary}}\delta_{s_i,s_j} + S^2_{\mathrm{shift}}, \quad (3)$$

$$S_3(i,j) = S_2(i,j) + S^3_{\mathrm{pairpot}}, \quad (4)$$

$$S_4(i,j) = -\big(F^{\mathrm{seq,0.001}}_{\mathrm{query}}(i)\cdot M^{\mathrm{str,org}}_{\mathrm{template}}(j) + F^{\mathrm{seq,0.001}}_{\mathrm{template}}(j)\cdot M^{\mathrm{seq}}_{\mathrm{query}}(i)\big)/2 - W^4_{\mathrm{2ndary}}\delta_{s_i,s_j} + S^4_{\mathrm{shift}}, \quad (5)$$

$$S_5(i,j) = S_4(i,j) + S^5_{\mathrm{pairpot}} \quad (6)$$

Here, $M^{\mathrm{str,org}}_{\mathrm{template}}(j)$ is the position-specific substitution matrix transformed from the original structure-derived probability profile $F^{\mathrm{struc,org}}_{\mathrm{template}}(j)$ by PSIBLAST utility programs (5). w^2_{2ndary}, w^4_{2ndary}, S^2_{shift}, S^4_{shift} are secondary structure profile weights and constant shift factors. S^3_{pairpot}, S^5_{pairpot} are pair potential terms that are constructed at stage $m = 2$ and $m = 4$ of PROSPECTOR_3 (55), respectively. All the weight and shift factors and gap penalties for dynamic alignment (58) are taken directly from PROSPECTOR_3 without further optimization because Eqs. 3–6 are similar to the corresponding ones in PROSPECTOR_3. Even so, our results show that PRO-SP^3 threading is already better than the consensus from the three state-of-the–art method PROSPECTOR_3, SP^3 and SPARKS for medium/hard targets (see Table 2).

Given a target sequence and a template structure, the threading approaches developed here may not be good enough to provide the best sequence-to-structure alignment for medium/hard targets. To remedy this shortcoming, we apply a parametric alignment method (25,59) to generate an ensemble of alignments and use TASSER-QA to select the ''best'' alignments. We perform this procedure by gridding the parameter space (w_0, w_1, w_{2ndary}, w_{struct}, S_{shift}) of the score $S_1(i,j)$ in Eq.1, where (w_0, w_1) are gap penalties. Each parameter is sampled with 0.0, 0.5, 1.0, 1.5, and 2.0 times the original values except for the gap opening penalty, w_0, which is sampled with 0.5, 1.0, 1.5, 2.0, and 2.5 times the original value. Distinct alignments are ranked according to their coverage and the top 1000 alignments kept. We then apply the TASSER-QA algorithm to select the top alignment model from these 1000 alignments. We call this selected top alignment the alternative alignment of the corresponding template.

Generation of the structure ensemble by multiple short TASSER simulations

A detailed flowchart of the model ensemble generation procedure is shown in Fig. 1 *B*. To generate an ensemble of diverse quality models, we apply a short TASSER refinement procedure on various sets of templates identified independently by the five PRO-SP^3 scores. Since the TASSER structure prediction method is described in detail in the literature (33,41,43,44,50,51,60), we merely present a brief overview of its essential components. Essential components of the TASSER force field are the set of side chain contact and the distance restraints derived from a set of input models. TASSER also requires a set of initial structures. Generally, the structures that provide the restraints and the starting conformations are the same. However, they can be different. For example, in chunk-TASSER (44), the set of structures used for deriving contact and distance restraints contains ab initio folded chunks in addition to threading templates, whereas the set for initial starting structures consists entirely of threading templates.

Since full TASSER modeling takes too long with large proteins, we use the same strategy that was employed in TASSER-lite (60) to limit the modeling procedure to short times (typically 10 hs) for all target sizes. The sets of template alignments used by TASSER or chunk-TASSER are:

a. Sets 1–5, each set consists of the top 10 template alignments from each of the five distinct threading approaches.
b. Set 6, the top two templates from each threading score, total of 10 templates.
c. Set 7, for medium/hard targets only, alternative alignments of the top two templates from each threading score, for total of 10 templates.
d. Set 8, the top four templates from each threading score, for a total of 20 templates.
e. Set 9, the top six templates from each threading score, for a total of thirty templates.
f. Set 10, the top 10 templates from each threading score, for a total of fifty templates.
g. Set 11, combined Set 6 and Set 7, for medium/hard targets only, for a total of 20 templates.

The differences between set d, e, and f are the number of templates from each threading score that are included. This can be used to explore the effects of different combinations and diversity of templates identified by the individual scoring functions.

We put the top five models from each short TASSER simulation into the structure ensemble. Each set provides from several to more than one hundred models depending on the scenario of TASSER simulation. TASSER uses modified

replica-exchange Monte Carlo for sampling (61). Each replica needs an initial structure. There are two kinds of scenarios in the TASSER simulations based on how the initial structures are assigned to the replicas. For a scenario A simulation, each template in the input set is taken as the initial structure of one replica. For a scenario B simulation, only one template in the input set is used as initial structure for all replicas. Therefore, for an input set of N templates, by shifting the template for the initial structure, one can apply N different scenario B simulations and one scenario A simulation. The models from a scenario A simulation are usually close to the average structure of all the inputs, whereas those from scenario B are close to the particular template assigned to the replicas. Here, for medium/hard targets, we apply scenario A simulations on Sets 1–10, and scenario A and B simulations on Set 11. For easy targets, only scenario A and B simulations on Set 8 are applied. The procedure generates up to 155 models (cluster centroids) for medium/hard targets and 105 models for easy targets. For medium/hard targets, the above procedure generates an ensemble that have models close to the top 10 template alignments and the top two alternative template alignments of each threading scoring function. For easy targets, the ensemble contains models close to the top four template alignments from each threading scoring function. Therefore, the ensemble has more diverse models than a single threading method can provide.

Full TASSER refinement on selected models and final model selection

After the generation of the ensemble of full-length structures that consists of top five cluster centroids provided by SPICKER (62) from all short TASSER simulations, we select the top 20 possible best quality models from the ensemble by the TASSER-QA (54) quality assessment method. These are then used to derive TASSER contact and distance restraints and also serve as the starting structures for full TASSER or chunk-TASSER refinement. One single scenario A simulation is applied to each target, and SPICKER is used again to obtain the top five cluster centroid structures for subsequent final model selection. To increase the chance of getting good final models, we select final models from both ensembles generated by short TASSER refinement and by full TASSER refinement. For target of ~100 residues, a full TASSER simulation takes ~20 h, but the simulation time increases rapidly with the increase of target size (roughly proportional to the square of size). A full simulation for a 300 residue target will take ~200 h.

RESULTS

To test our algorithm, we compiled a set of 723 benchmark proteins (whole chains or domain) structures that are all <250 residues in length. The threading library is built from structures all released before the testing benchmark structures. The testing dataset is classified by their SP^3 Z-score into 348 easy, 155 medium and 220 hard targets. The maximum sequence identity between target and library is 55%, and the average per target largest sequence identities between target and library for easy, medium and hard targets are 27%, 19%, and 18%, respectively. As mentioned in Wu et al. (16), comparison of methods is hard because of differences in the members of the template library. For example, if the library contains a close homolog, the results will be very different from the case where there is not even an analogous structure in the template library. Therefore, we compare pro-sp3-TASSER with our previous MetaTASSER algorithm that performed well in the CASP7 experiment (51). In CASP7 predictions by humans, the TASSER group results were based on the MetaTASSER server (51). Furthermore, the threading algorithms SP^3, SPARKS2 and PROSPECTOR_3 that provide inputs into MetaTASSER are among the best nonmeta threading methods (50,55,63). Therefore, MetaTASSER represents one of the best automated structure prediction methods in the literature. Since chunk-TASSER (44) shows a significant improvement over original TASSER for medium/hard targets, for fair comparison, we update MetaTASSER with chunk-TASSER for medium/hard targets. The target's TM-score to the native structure is used as structure quality measure in our comparison (53).

Table 1 compares the performance of pro-sp3-TASSER and MetaTASSER in terms of the number of targets having models with TM-score to their native structures of >0.4 and cumulative TM-scores, along with two-sided *p*-values of Student's *t*-test that show the statistical significance of the differences (64). Pro-sp3-TASSER has ≥10% more foldable targets than MetaTASSER for medium/hard target

TABLE 1 Comparison of pro-sp3-TASSER and MetaTASSER on the 723 protein benchmark set

	First model				Best of top five			
Target set (No. of targets)	MetaTASSER		Pro-sp3-TASSER		MetaTASSER		pro-sp3-TASSER	
Easy (348)	323	243.5	325	246.8(0.04)	326	248.8	332	254.2(6.9×10^{-6})
Medium (155)	60	58.7	62	59.2(0.55)	68	62.7	75	64.6(0.04)
Hard (220)	53	76.6	62	78.6(0.15)	78	85.5	90	88.0(0.01)
All (723)	436	378.8	449	384.6(0.01)	472	397.0	497	406.8(6.3×10^{-8})

Numbers in the first columns under each method are numbers of targets having first/best of top five models with TM-score to native >0.4. TM-score of 0.4 is a statistically significant threshold for structural similarity. The numbers in second columns are TM-scores. Numbers in parenthesis are two-sided *p*-values (64,65) between the two methods. A *p*-value of <0.05 is considered significant.

categories. For the first models, pro-sp3-TASSER shows on average a 1.4% TM-score improvement (p-value = 0.04) over MetaTASSER on the easy target set. In contrast, the improvements of TM-scores on medium/hard sets are not statistically significant (p-value = 0.55,0.15). For the best of top five models, pro-sp3-TASSER shows on average 2–3% improvements over MetaTASSER for all targets, with the improvements are statistically significant.

Although the percentage TM-score improvement is relatively small, it is a nontrivial achievement due to the fact that MetaTASSER plus chunk-TASSER is already a state-of-the-art method that performed well in CASP7 (51) and there are only a few methods that surpass it. The latest version of MetaTASSER has implemented some components of the methodology developed in this work, e.g., the special treatment for easy targets, selection of models from multiple TASSER simulations by TASSER-QA. However, it still does not match pro-sp3-TASSER's performance on medium/hard targets in the benchmark set. That indicates pro-sp3-TASSER is superior to MetaTASSER, particularly for medium/hard targets.

In Fig. 2, we show the target-to-target comparison of the two methods. Although, pro-sp3-TASSER does worse than MetaTASSER for certain targets, overall, pro-sp3-TASSER does better. The cases when pro-sp3-TASSER does worse are partly due to the failure to identify good templates. For example, for target 2ehp_A, although it is classified as easy target, only the SP^3 score component of PRO-SP^3 threading identifies the good template 1wqa_A, whereas in MetaTASSER, all three threading methods identify it. Another source of the poorer performance of pro-sp3-TASSER is that TASSER-QA fails to select good models. One such example is 2qwt_A. There are several models with TM-scores to native ~0.75–0.80 in the ensemble after the first round of TASSER modeling, but the TM-scores of the selected models are ~0.5–0.6. Therefore, the final best model by pro-sp3-TASSER only has a TM-score to native of 0.62, whereas the MetaTASSER model has a TM-score of 0.82. Fig. 3 shows the comparison of targets as a function of TM-score that demonstrate that pro-sp3-TASSER is better for all TM-score threshold cutoff values.

We next analyze the source of the improvement. In Table 2, the best of the top five templates or models that are inputs into TASSER are compared between MetaTASSER and pro-sp3-TASSER. Templates in MetaTASSER are ranked by the 3D-jury approach (28), whereas in the first round involving short TASSER simulations in pro-sp3-TASSER, the top five templates consist of the top one template from each threading score and in the second round involving the full TASSER simulation, they are ranked by TASSER-QA. For easy targets, PRO-SP^3 threading and MetaTASSER threading results are comparable. However, after TASSER-QA selection, the input model quality is significantly better. This is partly due to the fact that the selected models are full-length, whereas template models have unaligned regions. It is also reflects the refinement in the short TASSER simulations and the ability of TASSER-QA to select the best quality models. For medium/hard targets, templates from PRO-SP^3 threading are better overall than those from MetaTASSER threading. For a few targets, the TM-score of the best of top five model to native by MetaTASSER are ~0.3 whereas they are >0.5 from pro-sp3-TASSER (see Fig. 2). For these targets, only a single good template is identified and because MetaTASSER includes an overwhelming number of poor templates even though there is a single good template in the ensemble, the resulting good quality structures are not in the selected final top five models. Pro-sp3-TASSER is able to retain the template model by the special treatment mentioned earlier. For example, target 2yyv_A, the first template

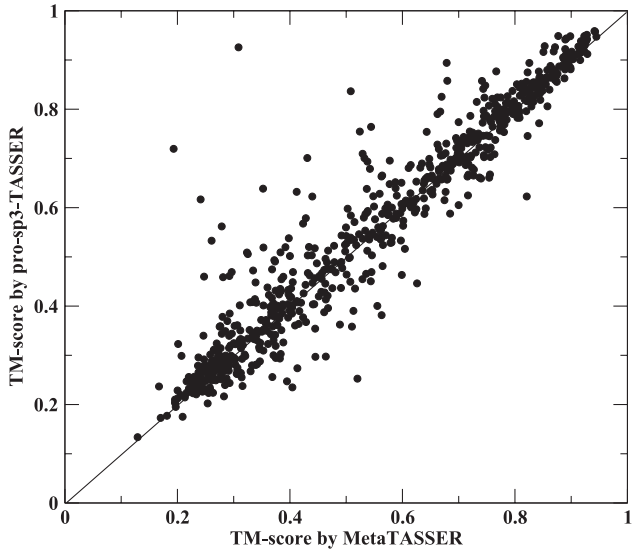

FIGURE 2 Scatter plot comparison between PRO-SP3-TASSER and MetaTASSER on the 723 benchmark set. TM-scores are from the best of top five models.

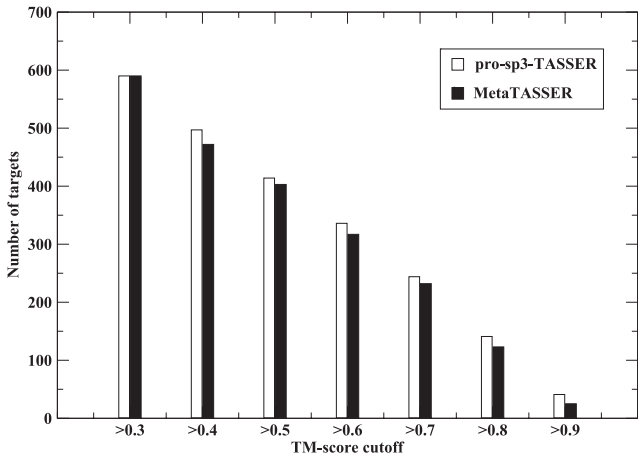

FIGURE 3 Histogram comparison between PRO-SP3-TASSER and MetaTASSER on the 723 benchmark set. TM-scores are from the best of top five models.

TABLE 2 Comparison between MetaTASSER and pro-sp3-TASSER of the cumulative TM-scores of the best of top five templates or models that are inputs into TASSER

Target set (No. of targets)	MetaTASSER Threading	PRO-SP^3 Threading	TASSER-QA selected
Easy (348)	238.1	238.1	247.4
Medium (155)	52.4	55.5	62.1
Hard (220)	67.7	70.8	84.7
All (723)	358.2	364.4	394.2

identified by all three of the MetaTASSER's threading algorithms is 1vr0_A, with a TM-score to native target of 0.89. However, because of the fact that the overwhelming majority of templates are of poor quality, the TM-score of the final best of top five models by MetaTASSER is only 0.31, whereas by pro-sp3-TASSER, the TM-score is 0.93.

Another source of improvement is the better template identification by PRO-SP^3 threading. One such example is 2uvp_A. In MetaTASSER, only the sixth template as ranked by SP^3 threading has a TM-score of 0.49 to the native structure, whereas all other templates have TM-scores ~0.2 to native. The preponderance of poor quality templates results in a TM-score of the final best model of 0.24. In pro-sp3-TASSER, there are five templates having TM-scores to native ranging from 0.44 to 0.64 and the final best TM-score is 0.62. An example that shows improvement by the TASSER-QA procedure in pro-sp3-TASSER is 2hz8_A, which is a 115 residue, four-helix bundle protein. The single best template from threading is 2b0h_A, which has a TM-score of 0.66 to native, is not ranked first by SP^3; therefore there is no special treatment in pro-sp3-TASSER. The alternative alignment model of template 2b0h_A has a better quality TM-score of 0.70 to native. After the first round of short TASSER modeling, TASSER-QA selects models with TM-scores ranging from 0.77 to 0.86 in the top 10 models. The final best of top five models given by pro-sp3-TASSER has a TM-score to native of 0.84, whereas the best model by MetaTASSER has 0.51. Thus, the initial models are refined by TASSER, and are then improved on in a subsequent TASSER iteration.

By limiting the total modeling time to 72 h for the entire pro-sp3-TASSER procedure, we implemented an automated server at http://cssb.biology.gatech.edu/skolnick/webservice/pro-sp3-TASSER/ for public use. The source code for standalone installation is also available upon request. For target sizes of ~100 residues, ~30 h are needed for modeling. However, for a target size of 300 residues, pro-sp3-TASSER will take ~200 h. Thus, the server is best suited for proteins <150 residues. For larger proteins, the limited time in the second round TASSER simulations may slightly degrade the final model quality.

DISCUSSION

We have shown that pro-sp3-TASSER shows considerable improvement of foldability (number of foldable targets) for medium/hard targets and on average, a statistically significant, 2–3% TM-score improvement over our previous MetaTASSER approach enhanced by chunk-TASSER. Improvements come from better template identification as well as better alignments and enrichment of better initial models by TASSER-QA selection from the large structure ensemble generated by short TASSER modeling. However, there is still much room for further improvements in the current approach. Fig. 4 shows the comparison of the TM-scores of the best models in the ensemble of structures (that contains up to 155 individual structures) generated by first round of short TASSER modeling and those of the best of top five models selected from the ensemble by TASSER-QA before full TASSER simulations. For many targets, especially for targets with models having TM-scores <0.7, the TM-score gap between the best possible model and the best of top five selected models is quite large. For all targets, the cumulative TM-score of best models in the ensemble is 426.9 compared to 394.2 for the selected best of top five models. The cumulative TM-scores of best models for subsets of easy, medium and hard targets are 254.4, 72.7, and 99.9, respectively. The corresponding values of selected models are 247.4, 62.1, and 84.7, respectively. The percentages of differences are 2.8%, 17.1% and 17.9% for easy, medium, and hard targets, respectively. Therefore, by developing a better model selection algorithm, one can considerably increase prediction accuracy, especially for medium/hard targets. For example, the PRO-SP3-TASSER predicted best model for 2ip6_A has a TM-score to native of 0.68, although there are, in fact, two models in the ensemble generated by short TASSER simulations that have TM-scores of 0.89 and 0.88. These models are ranked around 60th by TASSER-QA and originated from an

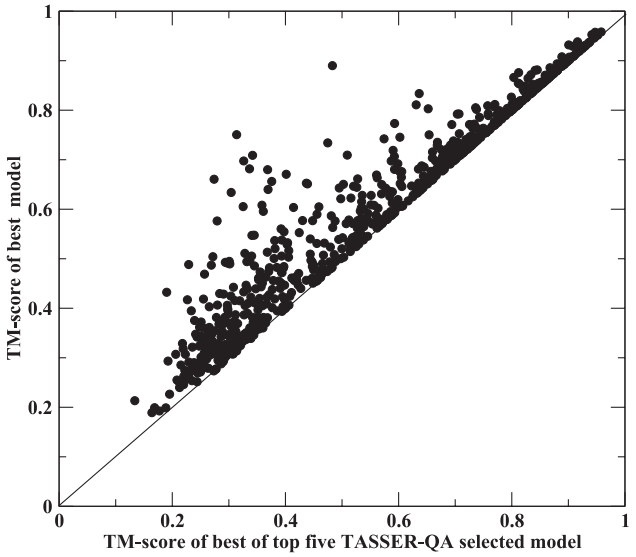

FIGURE 4 Comparison of TM-scores of best models in the ensemble of structures (up to 155) generated by first round of short TASSER modeling and those of the best of top five models selected from the ensemble by TASSER-QA before full TASSER simulations.

alternative alignment to template 2bl7_A (sequence identity 12%) with TM-score 0.73 (the SP^3 alignment has a TM-score of 0.59). Like the template identification problem for medium/hard targets, model quality assessment and selection is nontrivial. One direction for better model assessment/selection method is to use an all-atom physics-based force field, an approach that is currently under investigation in our laboratory.

This research was supported in part by grant Nos. GM-37408 and GM-48835 of the Division of General Medical Sciences of the National Institutes of Health.

REFERENCES

1. Baker, D., and A. Sali. 2001. Protein structure prediction and structural genomics. *Science.* 294:93–96.
2. Skolnick, J., J. Fetrow, and A. Kolinski. 2000. Structural genomics and its importance for gene function analysis. *Nat. Biotechnol.* 18:283–287.
3. Karplus, K., C. Barrett, and R. Hughey. 1998. Hidden markov models for detecting remote protein homologies. *Bioinformatics.* 14:846–856.
4. Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. A new approach to protein fold recognition. *Nature.* 358:86–89.
5. Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
6. Gribskov, M., A. D. McLachlan, and D. Eisenberg. 1987. Profile analysis: detection of distantly related proteins. *Proc. Natl. Acad. Sci. USA.* 84:4355–4358.
7. Jaroszewski, L., L. Rychlewski, W. Li, and A. Godzik. 2000. Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci.* 9:232–241.
8. Jaroszewski, L., L. Rychlewski, and A. Godzik. 2000. Improving the quality of twilight-zone alignment. *Protein Sci.* 9:1487–1496.
9. Yona, G., and M. Levitt. 2002. Within the twilight zone: a sensitive profile-profile comparison tool based on information theory. *J. Mol. Biol.* 315:1257–1275.
10. Marti-Renom, M. A., M. S. Madhusudhan, and A. Sali. 2004. Alignment of protein sequences by their profiles. *Protein Sci.* 13:1071–1087.
11. Mittelman, D., R. Sadreyev, and N. Grishin. 2003. Probabilistic scoring measures for profile-profile comparison yield more accurate short seed alignments. *Bioinformatics.* 19:1531–1539.
12. Fischer, D. 2000. Hybrid fold recognition: combining sequence derived properties with evolutionary information. *In* Pacific Symp. Biocomp. R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale, and T. E. Klein, editors. World Scientific, Hawaii. 119–130.
13. Ginalski, K., J. Pas, L. S. Wyrwicz, M. von Grotthuss, J. M. Bujnicki, et al. 2003. ORFeus: Dectection of distant homology using sequence profiles and predicted secondary structure. *Nucleic Acids Res.* 31:3804–3807.
14. Fischer, D., and D. Eisenberg. 1992. Protein fold recognition using sequence-derived predictions. *Protein Sci.* 5:947–955.
15. Liu, S., C. Zhang, S. Liang, and Y. Zhou. 2007. Fold recognition by concurrent use of solvent accessibility and residue depth. *Proteins.* 68:636–645.
16. Wu, S., and Y. Zhang. 2008. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins.* 72:547–556.
17. Shi, J., T. L. Blundell, and K. Mizuguchi. 2001. Sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* 310:243–257.
18. Zhou, H., and Y. Zhou. 2004. Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition. *Proteins.* 55:1005–1013.
19. Meller, J., and R. Elber. 2002. Protein recognition by sequence-to-structure fitness: Bridging efficiency and capacity of threading models. *Adv. Chem. Phys.* 120:77–130.
20. Teodorescu, O., T. Galor, J. Pillardy, and R. Elber. 2004. Enriching the sequence substitution matrix by structural information. *Proteins.* 54: 41–48.
21. Al-Lazikani, B., F. B. Sheinerman, and B. Honig. 2001. Combining multiple structure and sequence alignments to improve sequence detection and alignment: Application to the SH2 domains of Janus kinases. *Proc. Natl. Acad. Sci. USA.* 98:14796–14801.
22. Kelley, L. A., R. M. MacCallum, and M. J. E. Sternberg. 2000. Protein fold recognition using 1D and 3D sequence profiles coupled with secondary structure and solvation potential information. *J. Mol. Biol.* 299:501–522.
23. Kelley, L. A., R. M. MacCallum, and M. J. E. Sternberg. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* 299:499–520.
24. Zhou, H., and Y. Zhou. 2005. Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. *Proteins.* 58:321–328.
25. Chivian, D., and D. Baker. 2006. Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. *Nucleic Acids Res.* 34:e112.
26. Jones, D. T. 1999. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* 287: 797–815.
27. Cheng, J., and P. Baldi. 2006. A Machining learning information retrieval approach to protein fold recognition. *Bioinformatics.* 22:1456–1463.
28. Ginalski, K., A. Elofsson, D. Fischer, and L. Rychlewski. 2003. 3D-jury: a simple approach to improve protein structure predictions. *Bioinformatics.* 19:1015–1018.
29. Fischer, D. 2003. 3D-SHOTGUN: a novel, cooperative, fold-recognition meta-predictor. *Proteins.* 51:434–441.
30. Wu, S., and Y. Zhang. 2007. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res.* 35:3375–3382.
31. Cheng, J. 2008. A multi-template combination algorithm for protein comparative modeling. *BMC Struct. Biol.* 8:18–30.
32. Söding, J. 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 21:951–960.
33. Zhang, Y., and J. Skolnick. 2004. Automated structure prediction of weakly homologous proteins on genomic scale. *Proc. Natl. Acad. Sci. USA.* 101:7594–7599.
34. John, B., and A. Sali. 2003. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucl. Aci. Res.* 31:3982–3992.
35. Lee, S., and J. Skolnick. 2007. Development and benchmarking of TASSERiter for the iterative improvement of protein structure predictions. *Proteins.* 68:39–47.
36. Wu, S., J. Skolnick, and Y. Zhang. 2007. Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol.* 5:17–26.
37. Pillardy, J., C. Czaplewski, A. Liwo, J. Lee, D. R. Ripoll, et al. 2001. Recent improvements in prediction of protein structure by global optimization of a potential energy function. *Proc. Natl. Acad. Sci. USA.* 98:2329–2333.
38. Misura, K. M., D. Chivian, C. A. Rohl, D. E. Kim, and D. Baker. 2004. Physically realistic homology models built with ROSETTA can be more accurate than their templates. *Proc. Natl. Acad. Sci. USA.* 103:5361–5366.
39. Simons, K. T., C. Strauss, and D. Baker. 2001. Prospects for ab initio protein structural genomics. *J. Mol. Biol.* 306:1191–1199.
40. Bradley, P., K. Misura, and D. Baker. 2005. Toward high-resolution de novo structure prediction for small proteins. *Science.* 309:1868–1871.
41. Zhang, Y. 2007. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins.* 69 (*Suppl 8*):108–117.

42. Borreguero, J., and J. Skolnick. 2007. Benchmarking of TASSER in the ab initio limit. *Proteins.* In press.

43. Zhang, Y., A. Kolinski, and J. Skolnick. 2003. TOUCHSTONE II: A new approach to ab initio protein structure prediction. *Biophys. J.* 85:1145–1164.

44. Zhou, H., and J. Skolnick. 2007. Ab initio protein structure prediction using chunk-TASSER. *Biophys. J.* 93:1510–1518.

45. Odziej, S., C. Czaplewski, A. Liwo, M. Chinchio, M. Nanias, et al. 2005. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: Assessment in two blind tests. *Proc. Natl. Acad. Sci. USA.* 102:7547–7552.

46. Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. B, et al. 1977. The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures. *J. Mol. Biol.* 112:535–542.

47. Kihara, D., and J. Skolnick. 2003. The PDB is a covering set of small protein structures. *J. Mol. Biol.* 334:793–802.

48. Zhang, Y., I. Hubner, A. Arakaki, E. Shakhnovich, and J. Skolnick. 2006. On the origin and highly likely completeness of single-domain protein structures. *Proc. Natl. Acad. Sci. USA.* 103:2605–2610.

49. Zhang, Y., and J. Skolnick. 2004. The protein structure prediction problem could be solved using current PDB library. *Proc. Natl. Acad. Sci. USA.* 102:1029–1034.

50. Zhang, Y., A. Arakaki, and J. Skolnick. 2005. TASSER: An automated method for the prediction of protein tertiary structures in CASP6. *Proteins.* 61(S7):91–98.

51. Zhou, H., S.B. Pandit, S. Lee, J. Borreguerro, H. Chen, et al. 2007. Analysis of TASSER based CASP7 protein structure prediction results. *Proteins.* 69(Suppl. 8):90–97.

52. Lee, S., and J. Skolnick. 2008. Benchmarking of TASSER_2.0: An improved protein structure prediction algorithm with more accurate predicted contact restraints. *Biophys. J.* 95:1956–1964.

53. Zhang, Y., and J. Skolnick. 2004. A scoring function for the automated assessment of protein structure template quality. *Proteins.* 57:702–710.

54. Zhou, H., and J. Skolnick. 2007. Protein model quality assessment prediction by combining fragment comparisons and a consensus Cα contact potential. *Proteins.* 71:1211–1218.

55. Skolnick, J., D. Kihara, and Y. Zhang. 2004. Development and large scale benchmark testing of the PROSPECTOR 3.0 threading algorithm. *Proteins.* 56:502–518.

56. Jones, T. D. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292:195–202.

57. Domingues, F. S., P. Lackner, A. Andreeva, and M. J. Sippl. 2000. Structure-based evaluation of sequence comparison and fold recognition alignment accuracy. *J. Mol. Biol.* 297:1003–1013.

58. Vingron, M., and M. S. Waterman. 1994. Sequence alignment and penalty choice. Review of concepts, case studies and implications. *J. Mol. Biol.* 235:1–12.

59. Waterman, M. S., M. Eggert, and E. Lander. 1992. Parametric sequence comparisons. *Proc. Natl. Acad. Sci. USA.* 89:6090–6093.

60. Pandit, S. B., Y. Zhang, and J. Skolnick. 2006. TASSER-Lite: An automated tool for protein comparative modeling. *Biophys. J.* 91: 4180–4190.

61. Zhang, Y., D. Kihara, and J. Skolnick. 2002. Local energy landscape flattening: parallel hyperbolic Monte Carlo sampling of protein folding. *Proteins.* 48:192–201.

62. Zhang, Y., and J. Skolnick. 2004. SPICKER: a clustering approach to identify near-native protein fold. *J. Comput. Chem.* 25:865–871.

63. Zhou, H., and Y. Zhou. 2005. SPARKS 2 and SP3 servers in CASP 6. *Proteins.* 7 (*Suppl.*):152–156 (Supplement CASP issue).

64. Marti-Renom, M. A., M. S. Madhusudhan, A. Fiser, B. Rost, and A. Sali. 2002. Reliability of assessment of protein structure prediction methods. *Structure.* 10:435–440.

65. Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1989. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press.