

Template-based protein structure modeling using TASSER^{VMT}

Hongyi Zhou, and Jeffrey Skolnick*

Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30318

ABSTRACT

Template-based protein structure modeling is commonly used for protein structure prediction. Based on the observation that multiple template-based methods often perform better than single template-based methods, we further explore the use of a variable number of multiple templates for a given target in the latest variant of TASSER, TASSER^{VMT}. We first develop an algorithm that improves the target-template alignment for a given template. The improved alignment, called the SP³ alternative alignment, is generated by a parametric alignment method coupled with short TASSER refinement on models selected using knowledge-based scores. The refined top model is then structurally aligned to the template to produce the SP³ alternative alignment. Templates identified using SP³ threading are combined with the SP³ alternative and HHEARCH alignments to provide target alignments to each template. These template models are then grouped into sets containing a variable number of template/alignment combinations. For each set, we run short TASSER simulations to build full-length models. Then, the models from all sets of templates are pooled, and the top 20–50 models selected using FTCOM ranking method. These models are then subjected to a single longer TASSER refinement run for final prediction. We benchmarked our method by comparison with our previously developed approach, pro-sp³-TASSER, on a set with 874 easy and 318 hard targets. The average GDT-TS score improvements for the first model are 3.5 and 4.3% for easy and hard targets, respectively. When tested on the 112 CASP9 targets, our method improves the average GDT-TS scores as compared to pro-sp³-TASSER by 8.2 and 9.3% for the 80 easy and 32 hard targets, respectively. It also shows slightly better results than the top ranked CASP9 Zhang-Server, QUARK and HHpredA methods. The program is available for download at <http://cssb.biology.gatech.edu/>.

Proteins 2012; 80:352–361.
© 2011 Wiley Periodicals, Inc.

Key words: template-based modeling; threading; alignment; SP³; TASSER.

INTRODUCTION

Despite considerable efforts to develop accurate template free approaches, template-based protein structure modeling is still the only reliable protein structure modeling method.¹ Template-based modeling involves (1) template identification, usually by threading; (2) construction of the target-template alignment; (3) model generation based on this alignment; and (4) refinement of the template models with the goal of generating closer structures to the target native state than that provided by the template. The accuracy of the final model depends on the alignment accuracy between the template and target as well as the quality of the subsequent refinement. An ideal threading method should identify the best templates, namely, those with the best structural similarity to the native structure. In practice, this is not always the case.² Furthermore, the target-template alignment may not be optimal, especially for hard targets.³ Hard targets are those with poor quality threading alignments and good/poor quality structure alignments, respectively. Advances have been made in increasing the accuracy of template identification and threading alignment by going from single sequence alignments⁴ to profile-profile alignments,^{5–12} to machine learning,^{13,14} and meta-threading.^{6,15,16} Recent studies show that using multiple templates yields better models than those obtained from a single best template approach.^{17–22}

Results from the latest critical assessment of protein structure prediction (CASP9)²³ show that the top ranked automatic servers all used multiple template information in template-based modeling.²³ The Zhang-Server,²⁴ which performed well across all levels of target difficulty, employed a locally installed meta-server, LOMETs,²⁵ that uses more than eight individual threading methods for template identification/target alignment and TASSER¹⁷ for refinement. HHpredA,^{26,27} which performed well for easy targets, but moderately well for hard targets, used an updated method for hidden Markov profile generation and a neural network trained to filter distance restraints from multiple templates.²³ The final models are built by MODELLER²⁸ using filtered distance restraints. RaptorX,^{29,30} however, performed

Grant sponsor: NIH; Grant number: GM-48835, GM-37408.

*Correspondence to: Jeffrey Skolnick, Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology, Atlanta, GA 30318. E-mail: skolnick@gatech.edu

Received 28 June 2011; Revised 25 August 2011; Accepted 4 September 2011

Published online 14 September 2011 in Wiley Online Library (wileyonlinelibrary.com).

DOI: 10.1002/prot.23183

Table I
Summary of Major Differences Between TASSER^{VMT} and Pro-sp3-TASSER

	TASSER ^{VMT}	Pro-sp3-TASSER
Template identification	SP ³ threading	SP ³ and four other threading scores derived from PROSPECTOR_3 ⁴⁷
Alignments	SP ³ alternative and HHSEARCH for all targets	SP ³ default for easy SP ³ default and SP ³ alternative for hard targets
Alternative alignment generation procedure	SP ³ parametric FTCOM and DFIRE selection Short TASSER refinement	SP ³ parametric TASSER-QA selection
Number of template sets	120 for easy 80 for hard targets	1 for easy 11 for hard targets
Model selection from structure pool for final refinement	FTCOM	TASSER-QA

well for hard targets and slightly worse for easy targets. RaptorX^{29,30} and its variants RaptorX-MSA and RaptorX-Boost employed an outstanding alignment method that uses a boost tree to fit the alignment matching scores and a new method of generating multiple template-sequence alignments.²⁰ Most methods used a fixed number of multiple templates for a given target to produce the intermediate or final models. For example, standard TASSER¹⁷ uses 20, 30, or 50 templates as input for easy (those targets with good quality threading alignments), or hard targets (those having templates with poor quality threading alignments but often having good structural alignments). Pro-sp3-TASSER,²² that performed fairly well in CASP8³¹ and CASP9,²³ uses a few different sets with a fixed number of templates for generating TASSER inputs.

Obvious lessons learned from the top performing approaches in recent CASPs are that the use of multiple templates is an effective way of improving template-based modeling methods and that further improvement can be achieved by improving target-template alignments and filtering or selecting those templates that are optimally compatible with the target. For example, the outstanding performance of Zhang-Server and the template-based part of QUARK could be attributed partly to the consensus information from many threading methods that serve as a filter to remove poor quality templates and bad alignments, the inclusion of template fragments by segment threading,³² the predicted contacts by SVMSEQ,³³ and finally TASSER refinement. The use of better alignments and a neural-network trained filter for distance restraints in HHpredA makes it stand out among other multiple template-based methods, especially for easy targets.

In this work, we follow a similar scheme as in our previous pro-sp3-TASSER³¹ approach and use multiple sets of templates for a given target. In contrast to pro-sp3-TASSER that uses five different component threading scores, our new approach TASSER^{VMT} (TASSER with variable number of multiple templates) uses SP³ thread-

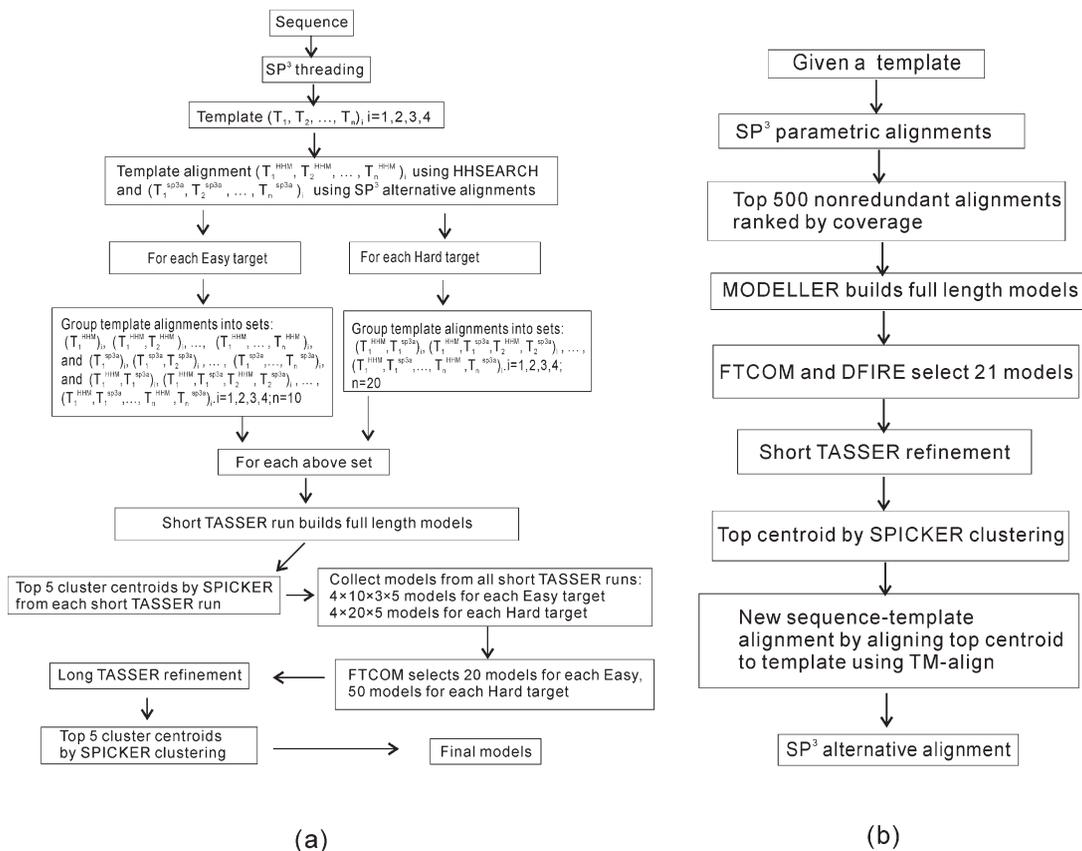
ing for template identification.³⁴ To build better models, we developed an algorithm to improve the target-template alignment for a given template. Alignments generated by HHSEARCH²⁶ are also included, since our tests show that they are slightly better than the default SP³ threading alignments for easy targets and could provide complementary sampling of alignment space. Then, short TASSER simulations build a pool of structures derived from different combinations of templates and their associated target alignments. The rationale of using subsets of a given multiple template set is to explore different combinations of templates to sample distance restraints or contacts to find the subset that are optimally compatible with the target. FCOM serves as an effective filter equivalent to the neural-network trained distance restraint filter in HHpredA to select the top 20–50 models that are then subjected to longer TASSER refinement. Table I summarizes the major methodological differences between current TASSER^{VMT} and previous pro-sp3-TASSER²² methods. We benchmarked TASSER^{VMT} against pro-sp3-TASSER²² on a set of 874 easy and 318 hard targets. We also tested TASSER^{VMT} on the 112 CASP9 targets and compare the results to top performing CASP9 methods, Zhang-Server, QUARK, HHpredA, RaptorX, Seok-server, MULTICOM_CLUSTER, and BAKER-ROSETTASERVER. We show that TASSER^{VMT} represents a promising approach to template-based modeling.

METHOD

The overall flowchart of TASSER^{VMT} is shown in Figure 1(a). Figure 1(b) shows the SP³ alternative alignment generation procedure that is part of Figure 1(a). In what follows, we shall describe the approach in more detail.

SP³ alternative alignment generation

For a given template, our goal is to improve the target-template alignment over that provided by SP³ threading. Targets are classified as easy if the SP³ threading

**Figure 1**

(a) Flowchart of TASSER^{VMT}. (b) Flowchart of SP³ alternative alignment generation.

Z-score of the top template is ≥ 6.0 , as medium if $4.5 < Z\text{-score} < 6.0$, and as hard if $Z\text{-score} \leq 4.5$.³⁴ We refer them as SP³ easy, medium, and hard. In practice, here we lump the medium and hard targets in the same category as hard for the prediction protocol. This classification is for prediction purposes and depending on predicted target difficulty allows for the use of different parameters in TASSER simulations and different template choices. (In contrast, in the results section, easy and hard targets are defined for assessment purposes using the actual model quality of the predicted models.) In *prop3-TASSER*,²² we generate alternative alignments only for SP³ hard targets using a parametric alignment generation method³⁵ and select the top alignment simply by TASSER-QA³⁶ without any refinement. Here, we also generate alternative alignments for easy targets and employ short TASSER refinement.

Given a target sequence and a template structure, we use a parametric alignment method^{35,37} to generate an ensemble of alignments. This is realized by a grid search on the five dimensional parameter space (w_0 , w_1 , $w_{2\text{ndary}}$, w_{struc} , S_{shift}) of the SP³ threading score,³⁴ where (w_0 , w_1) are gap penalties, $w_{2\text{ndary}}$ is the secondary structure score

weight, w_{struc} is weight of the structurally derived profile of the template, and S_{shift} is a shift parameter. 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. Because a gap-opening penalty of zero could lead to unrealistic, highly gapped alignments, this value of w_0 value is not sampled. Alignments are ranked according to their coverage (fraction of target residues aligned), with the top 500 distinct alignments kept. We then employ MODELLER²⁸ to build the full-length models and add the side-chain, and apply FTCOM ranking³⁸ and the DFIRE statistical potential³⁹ to select the top 21 models from these 500 models. The reason to use FTCOM and DFIRE is that they are not consensus-based scores so that they could potentially select a few good models. We apply the two independent components of FTCOM,³⁸ fragment comparison and template comparison, separately; each component selects seven models; we then use DFIRE to select another seven models. A short TASSER refinement run (limited to 5 h) is conducted using the selected 21 models as input and the top model from the top cluster centroid is selected by SPICKER⁴⁰

clustering on the TASSER trajectories. We then use TM-align⁴¹ to generate the structural alignment to the template. The resulting template alignment comprises the new target-template alignment that we term the “SP³ alternative alignment” of the corresponding target-template pair. We also employed the HHSEARCH alignment, as an independent alignment to the SP³ alternative alignment for each template because it is slightly better than the SP³ default alignment for easy targets, and it provides alignments that could potentially be complementary to the SP³ alternative alignments for both easy and hard targets.

Template identification and construction of a variable number of multiple template sets

In this work, we use SP³ threading to identify templates.³⁴ We use four different output scores to rank templates to explore different combinatorial space of templates: (1) the default SP³ rank score, that is, the raw threading score minus the reverse threading score; (2) the raw threading score; (3) the raw threading score/alignment length; (4) the raw threading score/target length. For each target, four sets of ranked templates are collected from the top (10 for easy, 20 for hard targets) ranked templates corresponding to the above four scores. We denote them as $(T_1, \dots, T_{10})_i$ and $(T_1, \dots, T_{20})_i$ with $i = 1, 2, 3,$ and 4 denoting the templates from the four scores for easy and hard targets, respectively. T_k^{HHM} and T_k^{sp3a} denote the template alignment of the k th ranked template T_k in each $i = 1, 2, 3,$ and 4 , using the HHSEARCH and the above generated SP³ alternative alignments, respectively.

To explore the combinatorial space of multiple templates, for easy targets, the following sets of template alignments are used in subsequent model building: (1) sets of $(T_1^{\text{HHM}})_i, (T_1^{\text{HHM}}, T_2^{\text{HHM}})_i, \dots, (T_1^{\text{HHM}}, T_2^{\text{HHM}}, \dots, T_{10}^{\text{HHM}})_i$; (2) sets of $(T_1^{\text{sp3a}})_i, (T_1^{\text{sp3a}}, T_2^{\text{sp3a}})_i, \dots, (T_1^{\text{sp3a}}, T_2^{\text{sp3a}}, \dots, T_{10}^{\text{sp3a}})_i$; (3) sets of $(T_1^{\text{HHM}}, T_1^{\text{sp3a}})_i, (T_1^{\text{HHM}}, T_1^{\text{sp3a}}, T_2^{\text{HHM}}, T_2^{\text{sp3a}})_i, \dots, (T_1^{\text{HHM}}, T_1^{\text{sp3a}}, T_2^{\text{HHM}}, T_2^{\text{sp3a}}, \dots, T_{10}^{\text{HHM}}, T_{10}^{\text{sp3a}})_i$ with $i = 1, 2, 3,$ and 4 . For hard targets, the template alignment sets are $(T_1^{\text{HHM}}, T_1^{\text{sp3a}})_i, (T_1^{\text{HHM}}, T_1^{\text{sp3a}}, T_2^{\text{HHM}}, T_2^{\text{sp3a}})_i, \dots, (T_1^{\text{HHM}}, T_1^{\text{sp3a}}, T_2^{\text{HHM}}, T_2^{\text{sp3a}}, \dots, T_{20}^{\text{HHM}}, T_{20}^{\text{sp3a}})_i$ with $i = 1, 2, 3,$ and 4 . In these combinations, only templates within the same i ($= 1, 2, 3,$ and 4) are combined. The total template alignment sets for an easy target is 120 and for a hard target is 80.

It should be noted that the above sets of template alignment represent only a small portion of all possible combinations. For 10 ranked templates and a given alignment method, the 10 sets we used consist of <1% of all possible 1023 combinations. However, we have found that among all sets with k (≤ 10) templates, the one we used is the most significant one because it contains the best-ranked k templates of the 10 templates.

Building models by short TASSER simulations

For each set of multiple templates (120 for easy and 80 for hard targets), we run short TASSER¹⁷ simulations (limited to 10 h to control the total modeling time) to build full length models. Each TASSER simulation contributes up to five models from the top five cluster centroids by SPICKER⁴⁰ clustering on the TASSER trajectories. The total number of models built for each easy target is 600 and 400 for each hard target. We found that for hard targets, building more models from the templates sets $(T_1^{\text{HHM}})_i, (T_1^{\text{HHM}}, T_2^{\text{HHM}})_i, \dots, (T_1^{\text{HHM}}, T_2^{\text{HHM}}, \dots, T_{10}^{\text{HHM}})_i, (T_1^{\text{sp3a}})_i, (T_1^{\text{sp3a}}, T_2^{\text{sp3a}})_i, \dots, (T_1^{\text{sp3a}}, T_2^{\text{sp3a}}, \dots, T_{10}^{\text{sp3a}})_i$ with $i = 1, 2, 3,$ and 4 does not give obvious improvement. Therefore, these sets of template alignments are not used for hard targets but are used for easy targets where an increase in the number of template sets did yield some improvement. The reason could be that for easy targets, good models dominate in the additional ensemble of models; whereas for hard targets, bad models dominate and lead to more false positives in the next step of model selection.

TASSER model refinement

After the pool of models is generated by short TASSER simulations, the top 20/50 models for each SP³ easy/hard target are selected by FTCOM,³⁸ a model assessment method that combines fragment comparison and template comparison scores. The selected models are subsequently input into TASSER for refinement. After around 20 h of simulation, TASSER trajectories were clustered with the SPICKER method⁴⁰ and top five cluster centroids were used for the final prediction. Possible clashes were removed and the main-chain, side-chain atoms built with Pulchra.⁴² The longer TASSER refinement simulation time is limited to 20 h based on the requirement of 72 h of total modeling time for servers as required in CASP.²³ However, if total modeling time is not an issue, these time restrictions can be relaxed and possibly better results might be obtained.

RESULTS

We tested our method on a large-scale benchmark set and compared the results with our previous method sp3-TASSER.^{22,31} For the CASP9 target set, we also compared our results to the top performing approaches of in CASP9.²³

Large-scale benchmark

A data set of 1192 targets was constructed from the SP³ threading library. These targets share a sequence identity <35% among themselves and to the threading template library used for template identification. All

Table II

Comparison of TASSER^{VMT} with Pro-sp3-TASSER on the 1192 Protein Benchmark Set

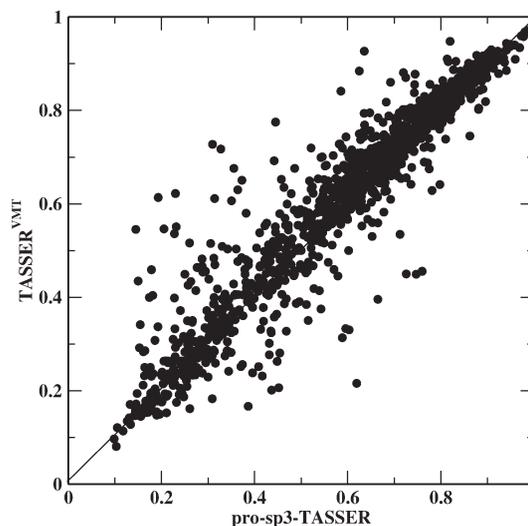
	874 Easy targets ^a		318 Hard targets	
	Cumulative GDT-TS	<i>P</i> -value ^b	Cumulative GDT-TS ^c	<i>P</i> -value
TASSER ^{VMT}	615.27	–	100.36 (85)	–
Pro-sp3-TASSER	594.62	7.7×10^{-20}	96.22 (61)	1.9×10^{-4}

^aA target is defined as easy if the TM-score to native of the first model by at least one of the two methods is higher than 0.5; otherwise, it is classified as hard.

^bTwo-sided *P*-value of the Student-t test between TASSER^{VMT} and pro-sp3-TASSER. A *P*-value of ≤ 0.05 is considered statistically significant.

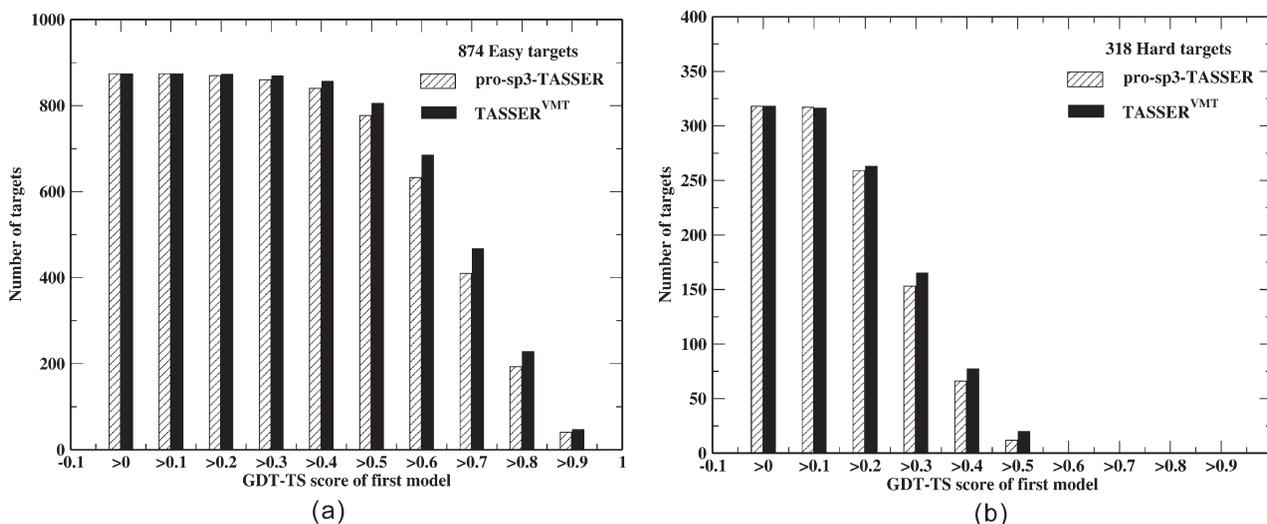
^cNumbers in brackets are the foldabilities defined as the number of targets having a first model TM-score to native ≥ 0.4 .

library structures were released before the targets. We compare TASSER^{VMT} with our previous developed method pro-sp3-TASSER using exactly the same template library. In our assessment, a target is classified as easy if the predicted first model has a TM-score⁴³ to native > 0.5 ; otherwise, it is defined as a hard target. The benchmark set has 874 easy and 318 hard targets, respectively. In accordance with convention, the first model's GDT-TS score⁴⁴ is employed for model quality assessment. In Table II, we compare the performance of TASSER^{VMT} and pro-sp3-TASSER as assessed by the cumulative GDT-TS scores of the first model as well as the foldability of hard targets. Foldability is defined as the number of targets whose first model has a TM-score to native ≥ 0.4 . Two structures having a TM-score ≥ 0.4 are considered to have significant structural similarity.^{43,45} On average, TASSER^{VMT} improves the GDT-TS scores by 3.5 and 4.3% for easy and hard targets, respectively. These

**Figure 2**

Scatter plot comparison of the first model's GDT-TS scores by pro-sp3-TASSER and TASSER^{VMT} on the 1192 protein benchmark target set. The number of targets for which TASSER^{VMT} is better/worse than pro-sp3-TASSER is 765/397.

improvements are statistically significant as shown by their respective *P*-values. For the 318 hard targets, foldability improves from 61 to 85 targets. Figure 2 shows the scatter plot of the comparison of GDT-TS scores for the two approaches. While the majority (765) of targets is improved by TASSER^{VMT}, there are also targets (397) that get worse relative to pro-sp3-TASSER. Figure 3(a,b) show the cumulative GDT-TS score histograms for

**Figure 3**

Histogram comparison of first model GDT-TS score by pro-sp3-TASSER and TASSER^{VMT} on the 1192 target set. (a) easy set; (b) hard set. The *y*-axis is the number of targets having first model GDT-TS score $> x$.

Table IIIAlignment Accuracy on the Top 10 SP³ Templates for the 1192 Benchmark Set

GDT-TS score to native ^a				
	Benchmark set		CASP9 set	
	874 Easy	318 Hard	80 Easy	32 Hard
	Change relative to SP ^{3b}	Change relative to SP ³	Change relative to SP ³	Change relative to SP ³
SP ³ alternative alignment	2.9%	2.6%	2.6%	3.2%
HHSEARCH alignment	0.7%	-2.9%	1.4%	-2.8%
TM-align alignment	15.3%	40.7%	16.7%	38.8%
	Comparison with TM-align			
	Correct match ^c	Correct match	Correct match	Correct match
SP ³ alignment	58.95/68.09/72.17	12.87/19.04/23.16	58.58/67.21/71.02	19.80/25.84/29.73
SP ³ alternative alignment	60.77/70.62/74.47	13.09/19.61/23.96	60.19/69.08/72.75	20.75/27.96/32.23
HHSEARCH alignment	59.70/68.67/72.56	12.01/17.97/21.88	60.04/68.57/72.50	18.61/24.63/28.84

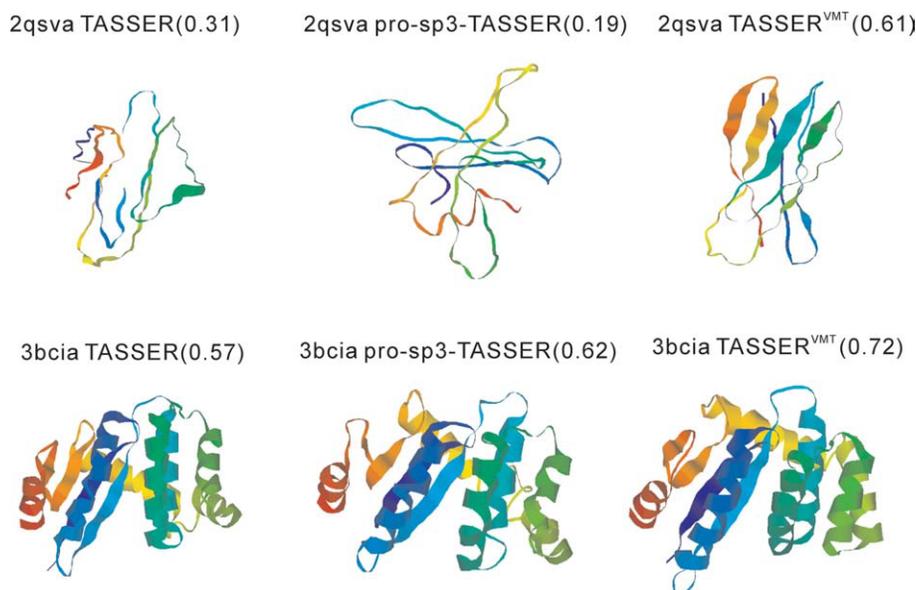
^aGDT-TS scores are calculated using full length models built from alignments with MODELLER²⁸.^bDefined as (model GDT-TS of this method-model GDT-TS of SP³)/(model GDT-TS of SP³).^cCorrect match is defined as that when the threading and structural alignments of the target residue to the template residue are identical, within ± 1 , or ± 2 (third number) residues apart. The number of matches in a given target is normalized by alignment length of TM-align.⁴¹ Presented numbers are the average per target per template and in percentages.

TASSER^{VMT} and pro-sp3-TASSER for easy and hard targets, respectively. TASSER^{VMT} improves over pro-sp3-TASSER models for all GDT-TS score cutoffs with the maximal improvements in GDT-TS score around 0.7 and 0.3 cutoffs for easy and hard targets, respectively.

We next analyze the alignment accuracy of our alternative alignment generation method in comparison with HHSEARCH and the structural alignment method TM-align⁴¹ using native structures. We shall compare the model quality as measured by GDT-TS scores to native of models built from the alignments using MODELLER²⁸ as well as directly compare alignments with the structural alignment using TM-align.⁴¹ In Table III, we show the average model GDT-TS score changes relative to using SP³ alignments and the average percentage of exact, and ± 1 , ± 2 shifted residue matches to the structural alignment as provided by TM-align for the top 10 SP³ templates. Apparently, both the SP³ alternative and HHSEARCH alignments are 1–3% better than SP³ alignments in terms of model GDT-TS scores to native for the easy targets. However, for hard targets, HHSEARCH is worse than SP³, whereas SP³ alternative alignments are still $\sim 3\%$ better than SP³. All methods are much worse than the best possible structural alignment generated by TM-align. For the exact match to the structural alignment given by TM-align, the SP³ alternative alignment is around 2% (absolute value) or 3% (relative value) better than the SP³ alignment for the easy targets. For the hard targets in the benchmark set, the average exact match to the structural alignment is quite low because many target-template alignments have zero matches. The relative improvement of SP³ alternative alignment (13.09%) over SP³ alignment (12.87%) for the hard targets is 1.7%, which is less than the GDT-TS score improvement of 2.6%. However, if the match is allowed to shift within ± 1 or ± 2 residues, then the relative improvement of SP³

alternative alignment over that of SP³ for the hard targets will be 3.0 or 3.5%, respectively. These improvements are consistent with the GDT-TS score result. Overall, the SP³ alternative alignment is consistently better than the SP³ alignment. A similar trend is also found for the CASP9 set described in the next subsection. Since the HHSEARCH provides an independent alignment for given target-template pair, we include it in our method to explore the alignment space more effectively.

In Figure 4, we present some examples of TASSER^{VMT} improvements over pro-sp3-TASSER. Target 2qsv_A domain 1 is a 101 residue hard target for threading method SP³. The predicted first model's GDT-TS score by pro-sp3-TASSER is only 0.19, whereas for TASSER^{VMT}, it is 0.61. The best model in the structure pool of pro-sp3-TASSER generated by short TASSER runs has a GDT-TS score of 0.60 while that of TASSER^{VMT} is 0.70 due to better alignments and more template combinations. The final prediction is dictated by the selected model quality for the TASSER refinement procedure. The GDT-TS score of the top model selected by FTCOM³⁸ in TASSER^{VMT} is 0.62, while that of the model selected by TASSER-QA³⁶ in pro-sp3-TASSER is 0.24. Therefore, the main reason for improvement is due to better model selection for this target. Another example is 3bci_A, a 165 residue easy target for SP³, whose GDT-TS improves from 0.62 to 0.72. Since it is an easy target, the model selection process alone does not make a big difference. The difference comes from the fact that the best model in the structure pool in pro-sp3-TASSER has a GDT-TS score of 0.63 while that from TASSER^{VMT} has a GDT-TS score of 0.73 due to better alignments employed. For comparison, TASSER models are also shown in Figure 4. Although on average, the pro-sp3-TASSER is better than TASSER,²² for the target 2qsv_A, it is slightly worse.

**Figure 4**

Examples of TASSER^{VMT} improvement over pro-sp3-TASSER models. Numbers in parenthesis are GDT-TS scores to native. TASSER models are also shown for comparison.

CASP9 targets

To compare TASSER^{VMT} with other state-of-the-art template-based methods, we tested TASSER^{VMT} on all 112 CASP9 targets. A target is defined as easy if the average TM-score⁴³ of the first models generated by the best 50% of all CASP9 servers is higher than 0.5; otherwise, it is defined as hard. The list of easy/hard targets can be found at <http://zhanglab.ccmb.med.umich.edu/casp9/>. This definition is equivalent to that in the above bench-

Table IV

Comparison of TASSER^{VMT} with Top CASP9 Servers on the 112 Targets

	80 Easy targets ^a		32 Hard targets	
	Cumulative GDT-TS	<i>P</i> -value ^b	Cumulative GDT-TS	<i>P</i> -value
TASSER ^{VMT}	54.28	–	10.57	–
Zhang-Server	53.50	0.04	10.91	0.20
QUARK	53.45	0.03	10.73	0.36
HHpredA	52.63	0.007	9.23	0.03
RaptorX	52.97	0.007	10.29	0.32
Seok-server	51.95	1.0×10^{-4}	9.91	0.12
MULTICOM_CLUSTER	51.72	2.9×10^{-5}	9.63	0.04
BAKER-ROSETTASERVER ^c	–	–	9.79	0.11
chunk-TASSER	50.43	2.0×10^{-11}	9.96	0.05
pro-sp3-TASSER	50.16	1.2×10^{-13}	9.67	0.01

^aA target is defined as easy, if the averaged TM-score of the first models by the best 50% of all CASP9 servers is higher than 0.5; otherwise, it is classified as hard. The list of targets can be found at <http://zhanglab.ccmb.med.umich.edu/casp9/>.

^bTwo-sided *P*-value of the Student-t test between TASSER^{VMT} and the given method. A *P*-value of ≤ 0.05 is considered statistically significant.

^cBecause of missing targets, we did not put results for easy targets here.

mark test where two methods are involved: TASSER^{VMT} and pro-sp3-TASSER. In what follows, CASP servers with small variations will be represented by one server. For example, HHpredA represents all three HHpred servers, RaptorX represents all RaptorX servers, and MULTICOM_CLUSTER represents all MULTICOM⁴⁶ servers.

To have a fair comparison, all necessary inputs (profiles, templates, fragments, secondary structure predictions, etc.) were generated using information that was available at the time of CASP9. The results are compiled in Table IV. TASSER^{VMT} outperforms all other methods for the 80 easy targets and the differences between TASSER^{VMT} and other methods are all statistically significant. TASSER^{VMT} is only slightly worse than Zhang-Server and QUARK for the 32 hard targets, although the differences are not statistically significant. This could be because the Zhang-Server and QUARK used template-free modeling for some hard targets. Compared to our own method, pro-sp3-TASSER, TASSER^{VMT} improves the results by 8.2 and 9.3% for easy and hard targets, respectively, and all improvements are statistically significant. The percentages of improvement are larger than those for the large benchmark set but the corresponding *P*-values are less significant. The difference is mainly due to different sample size (112 vs. 1192). For a small size sample set, for example, for the CASP9 set, a small number of outliers can dramatically influence the results.

To analyze the effects of using a variable number of template sets, the SP³ alternative alignment and HHSEARCH alignment, for this relatively small set of

Table VComparison of TASSER^{VMT} with Different Protocols on the 112 Targets

Protocol	SP ³ default alignment ^a	SP ³ alternative alignment ^a	HHSEARCH alignment ^a	Variable Multiple Templates ^a	80 Easy targets	32 Hard targets
					Cumulative GDT-TS	Cumulative GDT-TS
A ^b	+	–	–	–	49.34	9.27
B	+	–	–	+	51.59	9.95
C	–	+	–	+	52.06	10.17
D	+	–	+	+	53.12	9.96
E	+	+	–	+	53.48	10.45
TASSER ^{VMT}	–	+	+	+	54.28	10.57

^a+/- indicates the component is included/excluded in the protocol.^bIn this protocol, a single set of top 10/20 SP³ templates were used for easy/hard targets.

CASP targets, we conducted several different modeling protocols—A: Running TASSER refinements using a single set of top (10 for easy, 20 for hard targets, respectively) template models from SP³ without employing the variable multiple template protocol; B–E: protocols employing the Variable Multiple Template protocol with different combinations of SP³ default, SP³ alternative and HHSEARCH alignments. In protocol A, there is no intermediate stage of generating a structure pool. The results are compiled in Table V. Obviously, all other protocols are significantly better than protocol A due to the usage of a variable multiple template protocol. Thus, the variable multiple template protocol contributes a lot to TASSER^{VMT} performance. By comparing TASSER^{VMT} without SP³ alternative alignments (D in Table V, replacing SP³ alternative with SP³ default in TASSER^{VMT}) with the full TASSER^{VMT} protocol, we see the effect of the alternative alignments (in combination with the HHSEARCH alignment). For easy targets, the SP³ alternative alignments contribute around 2% of the increase in GDT-TS score; whereas for hard targets, this contribution is around 6%, which is higher than the average alignment accuracy increase given in Table III. This could be because the model selection procedure by FTCOM filtered out some worse alignments, while keeping the better ones.

The effect of SP³ alternative alignment without HHSEARCH can be obtained by comparing protocol C that uses only SP³ alternative alignments to B that uses only SP³ default alignments. Protocol C is better than B by 0.9/2.2% for easy/hard targets. Protocol E is the protocol of TASSER^{VMT} without HHSEARCH but with the combination of SP³ default and SP³ alternative align-

ments. It is equivalent to replacing the HHSEARCH alignment in TASSER^{VMT} with the SP³ default alignment. Protocol E is better than B by 3.6/5.0% and better than C by 2.7/2.8% for easy/hard targets. Thus, there is also combinatorial effect on the alignments.

The direct effect of HHSEARCH can be obtained by comparing protocol E with TASSER^{VMT}. Protocol E is slightly worse than TASSER^{VMT} by 1.5/1.1% for easy/hard targets. Therefore, the HHSEARCH alignment improves easy targets more than hard targets. HHSEARCH still improves hard targets even though on average the HHSEARCH alignment is worse than SP³ default for hard targets (see Table III). This could be due to fact that HHSEARCH is independent of SP³ alternative alignments and provides some complementary alignments, whereas SP³ default alignments and SP³ alternative alignments share common alignment profiles.

We now examine how good the SP³ parametric alignment for alternative alignment generation is. We compare the actual model quality of models built by MODELLER from the SP³ default alignment and from SP³ parametric alignments for the top 10 SP³ templates of the 112 CASP9 targets in Table VI. For each target-template pair, the average, standard deviation, best GDT-TS score of the models, and percentage of models having better GDT-TS scores than that of SP³ default alignment are obtained. The average model quality of the parametric alignments is slightly worse than that of the SP³ default alignment, which is understandable because the parameters of SP³ default alignment is optimized against structure alignments. The variance (standard deviation) of the parametric alignments is 11/14% of the default alignment

Table VIComparison of SP³ Default and SP³ Parametric Alignments on the 112 Targets^a

	SP ³ default		SP ³ parametric		
	GDT-TS	Average GDT-TS	Standard deviation of GDT-TS	Best GDT-TS	Percentage better than default
80 Easy targets	0.461	0.422	0.051	0.501	27.4%
32 Hard targets	0.236	0.210	0.032	0.279	31.2%

^aNumbers are the average per target per template for the top 10 SP³ templates. Models are built from the given alignment using MODELLER. The average, standard deviation, best, and percentage of better than default method are defined within models of the same target-template pair with variable SP³ parameters.

for easy/hard targets. Around 30% of the alignments are better than the default ones. The best alignment is on average 9/18% better than the SP³ default alignment for easy/hard targets. Thus, SP³ parametric alignment can generate models much better than the SP³ default alignment. Even though the best models are still far from the models generated by TM-align (see Table III; 16.7/38.8% better than SP³ default), they are much better than the SP³ alternative alignment (see Table III; 2.6/3.2% better than SP³ default). If a better model selection method could be used for alternative alignment generation, then the improvement would be more significant.

DISCUSSION

We have improved our automated template-based modeling approach pro-sp3-TASSER method significantly by improving alignment accuracy, more sampling of multiple template combinations and a better model selection method, FTCOM. The resulting method, TASSER^{VMT}, performs better than the best servers in CASP9 for easy targets. We note that in contrast to pro-sp3-TASSER and many other CASP servers, TASSER^{VMT}, does not separately consider any unaligned domains. However, based on our analysis of the two main components of our method, alignment and model selection, there is still much room for further improvement. As shown in Table III, current state-of-the-art alignment methods like HHSEARCH and the method developed in this work are far from optimal. Moreover, model selection could also be improved. For example, on average, the best model as measured by GDT-TS score in the structure pool generated by short TASSER simulations is 4.6 and 28% better than our predictions for easy and hard targets, respectively. A better model quality assessment prediction method could further improve these results, especially for hard targets. In summary, for easy/hard targets, improving the template alignment to coincide with the structural alignment would result in a ~15/40% improvement while improving model selection would provide another ~5/30% improvement. We are currently investigating the possibility of improving alignment and model selection further towards their upper limits. In this implementation of TASSER^{VMT}, we have explored only a relatively small number of multiple templates and generated alternative alignments by a simple parametric approach. It is possible that a more effective alternative alignment generation method that samples more near native alignments will result in further improvement. We are currently exploring a variety of approaches designed to achieve this goal.

ACKNOWLEDGMENTS

The authors thank Dr. Bartosz Ilkowski for managing the cluster on which this work was conducted.

REFERENCES

- Zhang Y. Progress and challenges in protein structure prediction. *Curr Opin Struct Biol* 2008;18:342–348.
- Venclovas C, Margelevicius M. Comparative modeling in CASP6 using consensus approach to template selection, sequence-structure alignment, and structure assessment. *Proteins* 2005;61:(Suppl 7):99–105.
- Xu J. Fold recognition by predicted alignment accuracy. *IEEE/ACM Trans Comput Biol Bioinfo* 2005;2:157–165.
- Fischer D, Eisenberg D. Protein fold recognition using sequence-derived predictions. *Protein Sci* 1992;5:947–955.
- Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* 1999;287:797–815.
- Fischer D. Hybrid fold recognition: combining sequence derived properties with evolutionary information. In: Altman RB, Dunker AK, Hunter L, Lauderdale K, Klein TE, editors. *Hawaii: World Scientific*; 2000. pp119–130.
- Jaroszewski L, Rychlewski L, Li W, Godzik A. Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci* 2000;9:232–241.
- Ginalski K, Pas J, Wyrwicz LS, von Grotthuss M, Bujnicki JM, Rychlewski L. ORFeus: detection of distant homology using sequence profiles and predicted secondary structure. *Nucleic Acids Res* 2003;31:3804–3807.
- Yona G, Levitt M. Within the twilight zone: a sensitive profile-profile comparison tool based on information theory. *J Mol Biol* 2002;315:1257–1275.
- Karplus K, Barrett C, Hughey R. Hidden markov models for detecting remote protein homologies. *Bioinformatics* 1998;14:846–856.
- Wu S, Zhang Y. MUSTER: improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins* 2008;72:547–556.
- Zhou H, Zhou Y. SPARKS2 and SP³ servers in CASP6. *Proteins* 2005;61(suppl 7):152–156.
- Lundström J, Rychlewski L, Bunnicki J, Elofsson A. Pcons: a neural-network-based consensus predictor that improves fold recognition. *Protein Sci* 2001;10:2354–2362.
- Xu Y, Xu D, Olman V. A practical method for interpretation of threading scores: an application of neural networks. *Statistica Sinica Special Issue on Bioinformatics* 2002;12:159–177.
- Fischer D. 3D-SHOTGUN: a novel, cooperative, fold-recognition meta-predictor. *Proteins* 2003;51:434–441.
- Ginalski K, Elofsson A, Fischer D, Rychlewski L. 3D-jury: a simple approach to improve protein structure predictions. *Bioinformatics* 2003;19:1015–1018.
- Zhang Y, Skolnick J. Automated structure prediction of weakly homologous proteins on genomic scale. *Proc Natl Acad Sci U S A* 2004;101:7594–7599.
- Zhang Y, Arakaki A, Skolnick J. TASSER: an automated method for the prediction of protein tertiary structures in CASP6. *Proteins* 2005;61(suppl. 7):91–98.
- Al-Lazikani B, Sheinerman FB, Honig B. Combining multiple structure and sequence alignments to improve sequence detection and alignment: application to the SH2 domains of Janus kinases. *Proc Natl Acad Sci U S A* 2001;98:14796–14801.
- Peng J, Xu J. A multiple-template approach to protein threading. *Proteins*, in press.
- Cheng J. A multi-template combination algorithm for protein comparative modeling. *BMC Struct Biol* 2008;8:18.
- Zhou H, Skolnick J. Protein structure prediction by pro-sp3-TASSER. *Biophys J* 2009;96:2119–2127.
- Moult J, Fidelis K, Kryshtafovych A, Tramontano A. 9th Critical Assessment of Techniques for Protein Structure Prediction. 2010. Available at: <http://predictioncenter.org/casp9/>.
- Zhang Y. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins* 2007;69(Suppl 8):108–117.
- Wu S, Zhang Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucl Aci Res* 2007;35:3375–3382.

26. Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* 2005;21:951–960.
27. Hildebrand A, Remmert M, Biegert A, Söding J. Fast and accurate automatic structure prediction with HHpred. *Proteins* 2009;77(Suppl 9):128–132.
28. Sali AaB TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815.
29. Peng J, Xu J. Boosting Protein Threading Accuracy. *Lect Notes Comput Sci* 2009;31:5541.
30. Peng J, Xu J. Low-homology protein threading. *Bioinformatics* 2010;26:i294–i300.
31. Zhou H, Pandit S, Skolnick J. Performance of the Pro-sp3-TASSER server in CASP8. *Proteins* 2009;77:123–127.
32. Wu S, Zhang Y. SEGMENTER: identifying protein sub-structural similarity by segmental threading. *Structure* 2010;18:858–867.
33. Wu S, Zhang Y. A comprehensive assessment of sequence-based and template-based methods for protein contact prediction. *Bioinformatics* 2008;24:924–931.
34. Zhou H, Zhou Y. Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. *Proteins* 2005;58:321–328.
35. Chivian D, Baker D. Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. *Nucl Aci Res* 2006;34:e112.
36. Zhou H, Skolnick J. Protein model quality assessment prediction by combining fragment comparisons and a consensus C α contact potential. *Proteins* 2007;71:1211–1218.
37. Waterman MS, Eggert M, Lander E. Parametric sequence comparisons. *Proc Natl Acad Sci U S A* 1992;89:6090–6093.
38. Zhou H, Skolnick J. Improving threading algorithms for remote homology modeling by combining fragment and template comparisons. *Proteins* 2010;78:2041–2048.
39. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci* 2002;11:2714–2726.
40. Zhang Y, Skolnick J. SPICKER: a clustering approach to identify near-native protein fold. *J Comput Chem* 2004;25:865–871.
41. Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucl Acids Res* 2005;33:2302–2309.
42. Rotkiewicz P, Skolnick J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J Comput Chem* 2008;29:1460–1465.
43. Zhang Y, Skolnick J. A scoring function for the automated assessment of protein structure template quality. *Proteins* 2004;57:702–710.
44. Zemla A, Venclovas C, Moulton J, Fidelis K. Processing and analysis of CASP3 protein structure predictions. *Proteins* 1999;3:22–29.
45. Zhang Y, Hubner I, Arakaki A, Shakhnovich E, Skolnick J. On the origin and highly likely completeness of single-domain protein structures. *Proc Natl Acad Sci U S A* 2006;103:2605–2610.
46. Cheng J, Wang Z, Tegge A, Eickholt J. Prediction of global and local quality of CASP8 models by MULTICOM series. *Proteins* 2009;77(Suppl 9):181–184.