# Chapter 3

# The MULTICOM Protein Tertiary Structure Prediction System

Jilong Li, Debswapna Bhattacharya, Renzhi Cao, Badri Adhikari, Xin Deng, Jesse Eickholt, and Jianlin Cheng

## Abstract

With the expansion of genomics and proteomics data aided by the rapid progress of next-generation sequencing technologies, computational prediction of protein three-dimensional structure is an essential part of modern structural genomics initiatives. Prediction of protein structure through understanding of the theories behind protein sequence–structure relationship, however, remains one of the most challenging problems in contemporary life sciences. Here, we describe MULTICOM, a multi-level combination technique, intended to predict moderate- to high-resolution structure of a protein through a novel approach of combining multiple sources of complementary information derived from the experimentally solved protein structures in the Protein Data Bank. The MULTICOM web server is freely available at http://sysbio.rnet.missouri.edu/multicom_toolbox/.

**Key words** Protein tertiary structure, Template recognition, Multiple template combination, Protein structure prediction, Structure quality evaluation, Structure quality enhancement

## 1 Introduction

The past few decades have witnessed an explosive growth in genomics and proteomics data. With the advancement of high-throughput genome sequencing technologies, the total number of gene and protein sequences is increasing exponentially. Therefore, in this genomic era, one vital goal for life scientists is to acquire knowledge from this vast repository of resources for better drug design and disease prevention strategies. Proteins fold into a three-dimensional structure, called tertiary structure, in order to carry out necessary biological functions, and therefore a high-resolution tertiary structure of a protein is the key to understanding and manipulating its biochemical and cellular functions. However, the rate of protein structure determination by experimental techniques (e.g., X-ray crystallography or NMR spectroscopy) lags far behind the rate of acquisition of new protein sequences primarily due to

Daisuke Kihara (ed.), *Protein Structure Prediction*, Methods in Molecular Biology, vol. 1137,
DOI 10.1007/978-1-4939-0366-5_3, © Springer Science+Business Media New York 2014

the time-consuming and expensive nature of the experimental methods. Therefore, the gap between known protein sequences and structure will continue to widen in the future making it impossible to experimentally solve the structures for all proteins. Consequently, less expensive and time-efficient computer-assisted prediction of protein tertiary structures is becoming increasingly popular.

Around 50 years ago, Anfinsen discovered the fact that all of the information necessary for RNase A to fold into its native structure is contained in its amino acid sequence, suggesting that the structure of a protein could be derived uniquely from its sequence alone [1]. Subsequently, interpretation of the sequence–structure relationships in proteins has become an active area of research in the field of biological sciences. As soon as the experimental structures of the first few proteins were made available, it became clear that evolutionarily related (homologous) proteins tend to retain the same overall three-dimensional fold (i.e., the arrangement and association of structural fragments) while accumulating some divergent mutations [2]. Moreover, despite being strongly correlated, structural divergence is much slower than sequence divergence [3]. These two important findings gave birth to one doctrine in protein structure prediction (also known as protein modeling) called homology modeling or comparative modeling (CM) [4]. Traditionally, this technique attempts to map the sequence of one protein (a target) to the sequence of another protein with a known structure (a template) to deduce the overall fold of the target and subsequently alter the target structure according to its sequence divergence with respect to the template. This approach is also commonly known as template-based modeling (TBM) and is one of the most widely used techniques in computational protein structure prediction. Intuitively, the success of TBM depends largely on the availability and ability to identify suitable templates for the target as well as the sequence similarity between the target and template. The accuracy is usually low when only a relatively distant homologous template is available for the target. Promisingly, constant efforts have been made by the community in the last decade, resulting in continual improvement of the accuracy of computationally based structure prediction.

With the aim of an objective assessment of the improvement in state-of-the-art methods for protein structure prediction, Moult and co-workers organized the biennial community-wide experiment called critical assessment of techniques for protein structure prediction (CASP) [5]. It was clear from the assessment of the CASP blind experiment that the accuracy of computational protein structure can be improved by combining information from multiple templates instead of relying on a single template [6–8]. This concept is at the heart of the MULTICOM protein structure prediction system [9]. MULTICOM essentially is a robust framework which aligns the target protein with multiple complementary templates and attempts to enhance the accuracy of structure prediction using a novel model combination approach followed by quality assessment techniques [10, 11] to refine the alternative models with the goal of selecting the best structure. MULTICOM officially made its debut in CASP8 [12], and the assessment of the results demonstrates the effectiveness of the method across diverse target difficulties (i.e., for easy cases where a suitable template can be identified to hard cases where only distantly homologous templates are available). With its consistent success during the CASP9 experiment, MULTICOM has been acknowledged by the community as one of the "best public CASP-certified protein structure prediction servers" (http://predictioncenter.org/index.cgi?page=links).

In the subsequent sections, we attempt to provide a thorough and comprehensive overview of the MULTICOM protein structure prediction suite. Subheading 2 (Materials) describes the input data, step-by-step instructions on how to use the MULTICOM web interface in order to generate the tertiary structure of a protein, and how to interpret the results. In Subheading 3 (Methods), we provide methodologies used to develop the multi-level combination pipeline used in MULTICOM. Two representative examples have been furnished in Subheading 4 (Case Studies) for users which describe the typical use of the system and the way to analyze the output. Subheading 6 (Notes) covers some beneficial tips to aid the users of MULTICOM on how to use the system seamlessly and resolve any potential issues during the execution of the pipeline or analysis of the results.

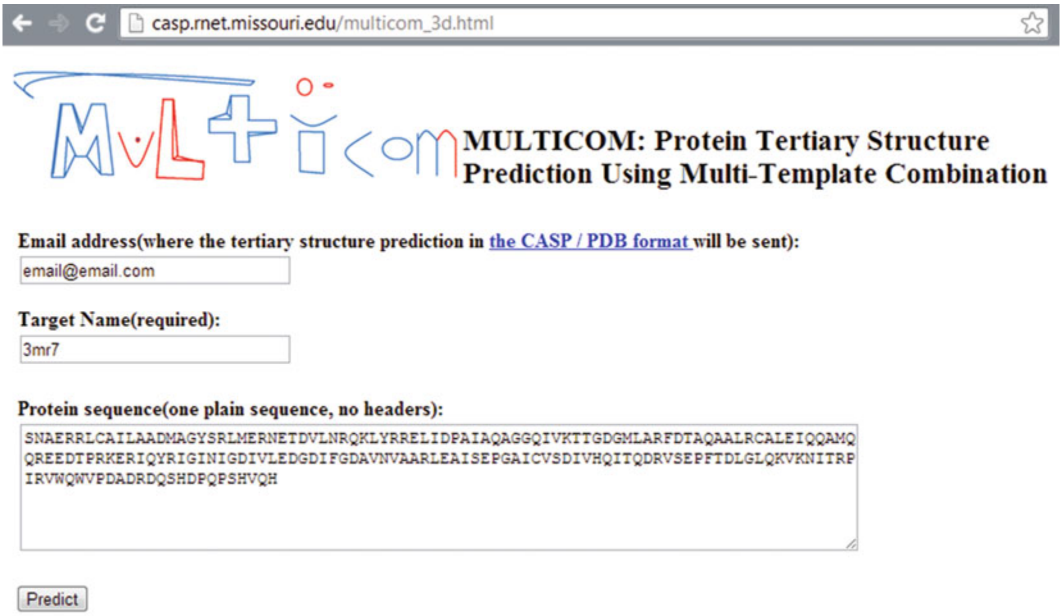
## 2 Materials

### *2.1 Input*

The input for the MULTICOM web server is the single-lettered amino acid sequence of the protein whose tertiary structure is to be predicted. The web server also needs a target name and e-mail address along with the amino acid sequence. The target name uniquely identifies the job, which is helpful when there is more than one job being submitted. The e-mail address is where the server sends the predicted model once the prediction is complete.

### *2.2 Usage*

Predicting a protein's structure using MULTICOM is a two-step process. The first step is to submit the amino acid sequence to the server and then wait for the results. The second step begins after the MULTICOM web server sends an e-mail with the predicted structure as an attachment. The attached structure file is a standard protein data bank (pdb) file and can be visualized, analyzed, or evaluated using any available tools.



MULTICOM: Protein Tertiary Structure Prediction Using Multi-Template Combination

Email address(where the tertiary structure prediction in the CASP / PDB format will be sent):

email@email.com

Target Name(required):

3mr7

Protein sequence(one plain sequence, no headers):

```
SNAERRLCAILAADMAGYSRLMERNETDVLNRQKLYRRELIDPAIAQAGGQIVKTTGDGMLARFDTAQAALRCALEIQQAMQ
QREEDTPRKERIQYRIGINIGDIVLEDGDIFGDAVNVAARLEAISEPGAICVSDIVHQITQDRVSEPFTDLGLQKVKNITRP
IRVWQWVPDADRDQSHDPQPSHVQH
```

**Fig. 1** The MULTICOM web server input page being filled with the sequence of chain A of a protein with PDB ID 3MR7. The input sequence is text wrapped in the text area and does not contain any white space characters

### *2.2.1 Step 1: Submit the Sequence*

The input sequence of amino acids should not contain any letters or characters other than the 20 standard amino acid symbols. Any special characters such as *, $, and & should be removed from the sequence. White space characters including space, newline, tab, and carriage return should also be removed from the sequence. Once the e-mail address, target name, and sequence fields are filled, clicking on the predict button displays a status page. Figure 1 shows an example input for chain A of the protein with PDB ID 3MR7. All data in the input fields, including the e-mail address, needs to be verified before clicking on the predict button.

### *2.2.2 Step 2: Download the Prediction*

Once the server completes the prediction, the results are sent to the corresponding e-mail address. The e-mail sent by the MULTICOM web server contains two attachments: model.pdb and align.pir. The PDB codes of the template sequences along with their alignment score are also included in the e-mail body as a list.

## *2.3 Output*

The pdb file attached is the standard pdb file that has the *x*, *y*, and *z* coordinates of each atom in the protein and is in standard CASP format(http://predictioncenter.org/casp8/index.cgi?page=format). The pir file attached is a multiple sequence alignment file that shows sequence alignment of the input sequence with the templates found during the prediction process and is used to generate the predicted structure. The pdb file can be visualized using any viewer tools such as Chimera [13], PyMOL [14], Rasmol [15], and Jmol [16].

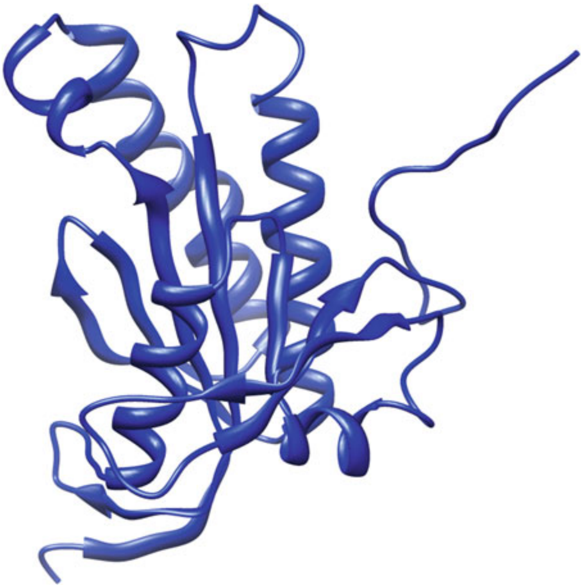


**Fig. 2** The MULTICOM web server's prediction for chain A of a protein with PDB ID 3MR7 visualized using PyMOL

Figure 2 shows an example of visualizing the model.pdb file predicted for chain A of a protein with PDB ID 3MR7. In case the native structure is also available, tools like TM-score [17] may be used to evaluate the prediction. Additionally, the alignment file may be analyzed for alignment information in order to understand the contribution of each template to the predicted model.

### 2.4 Availability

The MULTICOM web server is freely accessible at http://casp.rnet.missouri.edu/multicom_3d.html which is in the MULTICOM toolbox (http://sysbio.rnet.missouri.edu/multicom_toolbox/). Prediction time depends on factors including server load, length of the input sequence, and difficulty of the query (i.e., whether or not good templates can be found).

## 3 Methods

As shown in Fig. 3, there are five steps in the MULTICOM protein structure prediction system [9, 18]. The first step generates a number of templates and their sequence alignments for an input query sequence. The second step generates a number of query-template alignments. The third step creates several structures (also called protein models) for the query. The fourth step evaluates the quality of the generated models. The last step improves the quality of the generated models. Finally, the system outputs the predicted model with the best quality.

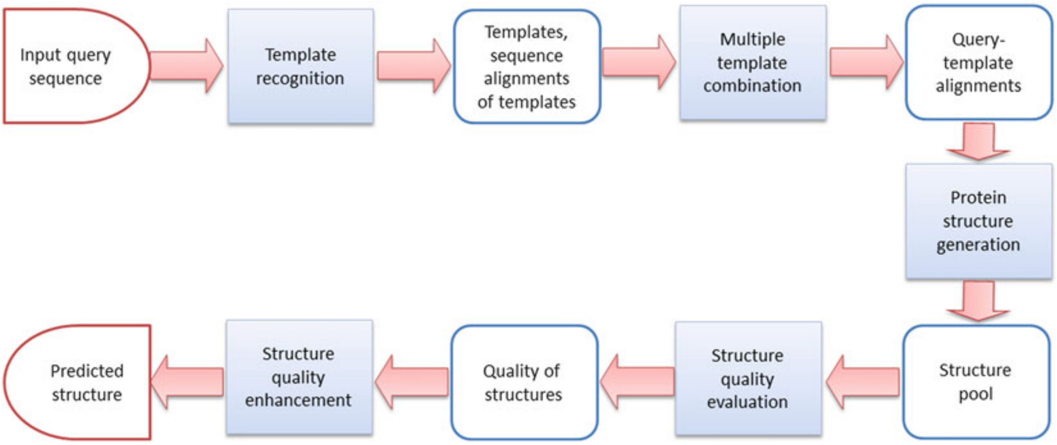

Input query sequence
Template recognition
Templates, sequence alignments of templates
Multiple template combination
Query-template alignments
Protein structure generation
Structure pool
Structure quality evaluation
Quality of structures
Structure quality enhancement
Predicted structure

**Fig. 3** The MULTICOM protein tertiary structure prediction system

### *3.1 Template Recognition*

Template recognition needs a template library in order to identify the templates for the query sequence. In this system, the template library has been constructed based on the PDB [19]. The template library includes information such as template sequence, template structure, secondary structure, solvent accessibility, and template sequence profiles.

In this step, sequences homologous to the query are first found by searching the query sequence against the non-redundant protein sequence database via PSI-BLAST [20]. The query and its homologous sequences are then searched against the template library by different search tools [20–27] in order to find a number of templates with information about the structure of the query. A number of templates with low *e*-values are generated after these searches, along with local alignments between the query and its templates (*see* **Note 1**). The top-ranked templates and their query-template alignments for each tool are saved separately. A consensus list of the top-ranked templates is also generated according to the number of times it is identified by each search tool.

### *3.2 Multiple Template Combination*

This step integrates multiple template structures coming from the previous step and generates a number of combined query-template alignments. This is done because multiple structurally similar templates may provide more accurate structural information for the query than a single template [6]. Three multiple template combination methods are used in this step.

The first method creates a combined query-template alignment based on the query-template alignments generated by each search tool. The combined query-template alignment contains the best query-template alignment and some other query-template alignments that have similar *e*-values with the best alignment. The aligned regions of all alignments have consistent structures (*see* **Note 2**).

The second method creates a combined query-template alignment based on the consensus list of templates. For each template, TM-Align [28] is used to align it with all other templates and the aligned regions are used to generate the multiple sequence alignment of this template. Then the multiple sequence alignment tool is used to align the multiple sequence alignments of all templates and that of the query to get the combined query-template alignment.

The third method uses three kinds of query-template alignments generated by PSI-BLAST [20], HHSearch [25], and SPEM [29] separately. This method combines these alignments for one query in this order: the PSI-BLAST local alignment, HHSearch alignment, and SPEM global alignment.

### 3.3 Protein Structure Generation

This step first checks the templates identified by the previous steps. If there are one or more templates which can cover the whole query or most of the query with very short unaligned regions (*see* **Notes 2** and **3**), the TBM tool Modeller [30] is used to generate a number of models. If there are no homologous templates or only one template covering a part of the query, a recursive protein modeling method [31] is used to generate the models. This method first uses the TBM tool Modeller [30] to model the regions which are aligned and covered very well by templates. We call these regions certain regions, while the unaligned regions are termed uncertain regions. A variant of Rosetta [31, 32] is used to construct other uncertain regions. Depending on the amount of template information available, the method may use only the TBM method or template-free modeling method or combine TBM method and template-free modeling method to generate a structure for the query. The final product of this step is a model pool for the query.

### 3.4 Structure Quality Evaluation

This step evaluates the quality of each model without knowing the native structure. In order to evaluate the quality of each model and identify the more accurate models, three structure quality evaluation methods are used. The first method (ModelEvaluator [33]) provides each model with an absolute quality score based on the features of that model (*see* **Note 4**). The secondary structure, solvent accessibility, contact map, and beta-sheet topology of the model can be parsed from the model directly, and they also can be predicted from the target sequence [34–36]. For each of them, we use the difference between that parsed from the model and that predicted from the target sequence as a feature. The second approach uses the structure alignment tool TM-score [17] to calculate the similarity score between the model and all other models in the model pool and then uses the average similarity score as the quality score of this model (*see* **Note 5**). The third method tries to combine the first two approaches. It selects the top models based on the quality score using the first method as the reference model set. Each model is compared with all models in the reference model set, and the average similarity score is used as the quality score. The local quality score of each residue is also calculated in this step. This is accomplished by aligning a model with each model in the reference model set. The distance between each residue in this model and its counterpart in a reference model in the reference set is calculated separately as a local quality score. Finally, the local quality score of each residue is the average distance of this residue and all of its counterparts.

### 3.5 Structure Quality Enhancement

In this step, the top-ranked models based on the structure quality evaluation are searched against the model pool to check if there exist other similar models (*see* **Note 6**). If there are some similar models, this step combines the top-ranked models with the similar models. Otherwise, very similar local regions of other models are combined with the top-ranked models. This model combination can usually get better models than the original top-ranked models. Moreover, the local quality score is also used for the structure quality enhancement. The regions with very poor local quality scores are resampled by a variant of Rosetta [31, 32] which constrains the local region modeling without changing other regions. The final prediction of this system is the best refined model.

## 4 Case Studies

As case studies, the MULTICOM web server was used to predict tertiary structure of the first chains (chain A) of two proteins: adenylate/guanylate cyclase/hydrolase from *Silicibacter pomeroyi* and diguanylate cyclase from *Pelobacter carbinolicus*. These proteins were also listed as prediction targets in CASP9 with target id as T0520 (http://predictioncenter.org/casp9/target.cgi?id=21&view=all) and T0634 (http://predictioncenter.org/casp9/target.cgi?id=178&view=all), respectively. These two protein sequences were supplied to the MULTICOM web server. The predictions were visualized using PyMOL and evaluated using TM-score and RMSD (average root mean square distance between the corresponding atoms) (*see* **Note 7**). The case studies show that the predicted structures are highly accurate with TM-score value of 0.9454 for target T0520 and 0.8547 for target T0634 and an RMSD value of 0.581 for T0520 and 1.257 for T0634. MULTICOM was ranked among the top ten predictors for both of these targets.

### 4.1 Case Study I

To predict the tertiary structure of adenylate/guanylate cyclase/hydrolase (from *Silicibacter pomeroyi*), its corresponding fasta sequence file was downloaded from PDB [19]. The PDB ID for

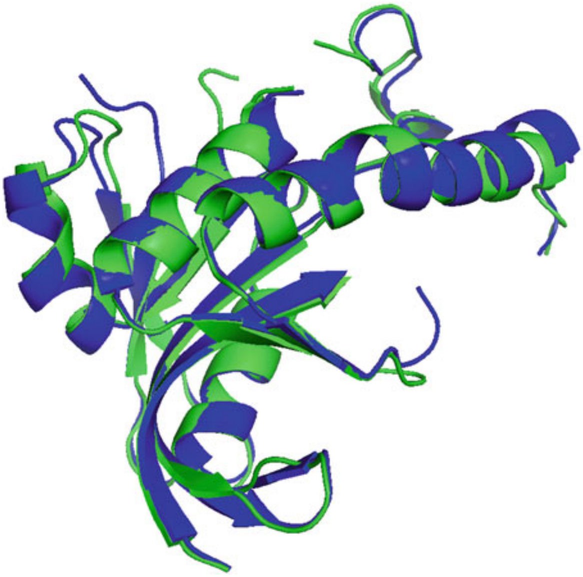


**Fig. 4** Filtered native structure (shown in *green* color) and MULTICOM-predicted filtered structure (shown in *blue* color) superimposed using PyMOL for protein adenylate/guanylate cyclase/hydrolase

this protein is 3MR7, and the fasta sequence file is available at http://www.rcsb.org/pdb/files/fasta.txt?structureIdList=3MR7. The sequence for chain A was copied to a separate text file to remove newline characters. After removing newline characters, the whole sequence, 189 characters long, was now in a single line that begins with the residues SNAE and ends with residues HVQH. The sequence was then copied and supplied as input to the MULTICOM web server as shown in Fig. 1. The server took 17 min to complete the task. The predicted structure (model.pdb) was then visualized with PyMOL. To visually compare the predicted structure with the native structure, the native structure was downloaded from http://www.rcsb.org/pdb/files/3MR7.pdb. Before performing the comparison, the native structure and predicted structure both need to be filtered for two reasons: the native structure has three chains, and the predicted structure has only one; thus, there may be disordered regions in predicted or native structures. Finally, the filtered predicted structure and filtered native structure were both superimposed and visualized in PyMOL as shown in Fig. 4. Additionally, the predicted structures were evaluated using TM-score and RMSD (*see* **Note** 7). The TM-score value of 0.9454 and RMSD value of 0.581 show that the prediction is very accurate.

### 4.2 Case Study II

To predict the structure of diguanylate cyclase (from *Pelobacter carbinolicus*), steps similar to Case Study I were executed. The PDB ID for this protein is 3N53, and the fasta file was downloaded from

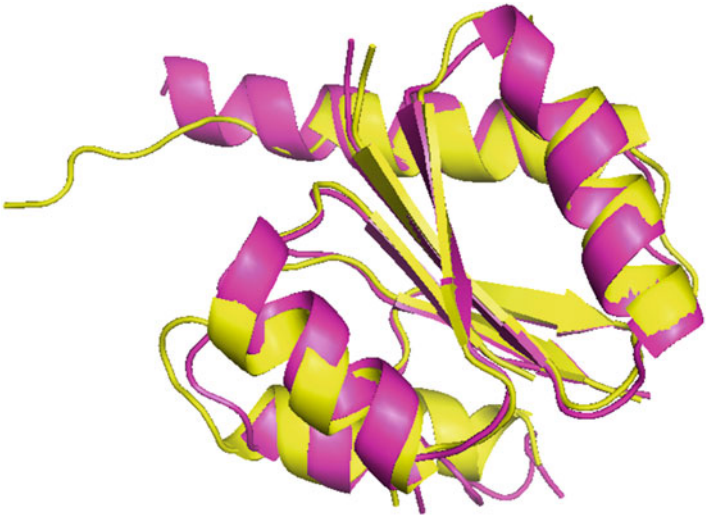


**Fig. 5** Filtered native structure (shown in *pink* color) and MULTICOM-predicted filtered structure (shown in *yellow* color) superimposed using PyMOL for protein diguanylate cyclase

http://www.rcsb.org/pdb/files/fasta.txt?structureIdList=3N53. The 140-residue-long sequence starting with MSLK and ending with HHHH was supplied to the web server, and it took around half an hour for the prediction to complete. Similar to Case Study I, the filtered predicted structure and filtered native structure were obtained, superimposed, and visualized in PyMOL as shown in Fig. 5. For this target as well, a high TM-score value of 0.8547 and RMSD value of 1.257 imply an accurate prediction.

## 5 Conclusion

Given the implications of protein structure in protein functional analysis and rational drug design as well as the limitations of existing experimental techniques to determine protein structure, computational approaches to predict protein structure will continue to be a necessity. The MULTICOM protein structure prediction pipeline stands ready to meet the needs of the research community and is accessible via a web service. The method uses a multi-level combination technique to combine multiple protein structure templates and sources of structural information to generate models and then employs a number of model refinement and selection tools to return the best possible predicted structure. The MULTICOM system is capable of using both template-based and template-free modeling to handle the full spectrum of protein modeling and generate predictions for all protein structure prediction tasks from the relatively easy to difficult. The system has been thoroughly and successfully tested in CASP8 and CASP9 and assessed as one of the best public, CASP-certified protein structure prediction servers.

## 6 Notes

1. An *e*-value is generated when using a search tool like BLAST [20, 21] to search the query against the template library. Usually, a low *e*-value means that the template has high similarity to the query.
2. Regions of a protein model usually refer to continuous segments of amino acids. Two regions have consistent structures if the similarity score between them is higher than a set threshold. The similarity score is calculated using the GDT-TS score generated from TM-score [17] when comparing them. In the MULTICOM system, we set the threshold to 0.75 for comparison of two regions.
3. Very short unaligned regions mean that there are less than ten residues unaligned in the template.
4. The absolute quality score of the model is the GDT-TS score between this model and its native structure. The GDT-TS score describes the expected similarity between the model and the native structure.
5. This approach is very sensitive about the input model pool. When the input model pool is small or contains many poor models, this approach does not work very well.
6. Two models are similar if the pairwise GDT-TS score is higher than a threshold. MULTICOM uses a threshold of 0.7 for comparison of two models.
7. TM-score [17], RMSD (average root mean square distance between the corresponding atoms), and GDT-TS score are commonly used tools to compare and evaluate protein structure predictions. The online version of the TM-score tool is available at http://zhanglab.ccmb.med.umich.edu/TM-score/. To compare the native structure (e.g., native.pdb) with a predicted structure (e.g., predicted.pdb), the predicted.pdb file is uploaded as Structure 1 and native.pdb is uploaded as Structure 2, leaving the e-mail address field blank. After running the comparison, the assessment results page shows the TM-score value.

## Acknowledgment

The work was supported by an NIH grant R01GM093123 to J.C.

## References

1. Anfinsen CB, Haber E, Sela M, White F Jr (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci USA 47(9):1309
2. Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. EMBO J 5(4):823
3. Bujnicki JM (2005) Protein-structure prediction by recombination of fragments. Chembiochem 7(1):19–27
4. Krieger E, Nabuurs SB, Vriend G (2003) Homology modeling. Methods Biochem Anal 44:509–524
5. Moult J, Pedersen JT, Judson R, Fidelis K (2004) A large-scale experiment to assess protein structure prediction methods. Proteins 23(3):ii–v
6. Cheng J (2008) A multi-template combination algorithm for protein comparative modeling. BMC Struct Biol 8(1):18
7. Fischer D (2003) 3D-SHOTGUN: a novel, cooperative, fold-recognition meta-predictor. Proteins Struct Funct Bioinf 51(3):434–441
8. Sali A, Blundell T (1994) Comparative protein modelling by satisfaction of spatial restraints. Proteins 64:C86
9. Wang Z, Eickholt J, Cheng J (2010) MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8. Bioinformatics 26(7): 882–888
10. Cheng J, Wang Z, Tegge AN, Eickholt J (2009) Prediction of global and local quality of CASP8 models by MULTICOM series. Proteins 77(S9):181–184
11. Wang Z, Tegge AN, Cheng J (2008) Evaluating the absolute quality of a single protein model using structural features and support vector machines. Proteins 75(3):638–647
12. Moult J, Fidelis K, Kryshtafovych A, Rost B, Tramontano A (2009) Critical assessment of methods of protein structure prediction: Round VIII. Proteins 77(S9):1–4
13. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera: a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
14. LLC DS The PyMOL molecular graphics system. http://pymol.sourceforge.net/
15. Sayle R (1994) RasMol v2. 5-Molecular visualisation program
16. Jmol: an open-source Java viewer for chemical structures in 3D. http://jmol.sourceforge.net/. Accessed 10 Dec 2008
17. Zhang Y, Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. Proteins 57(4):702–710
18. Li J, Deng X, Eickholt J, Cheng J (2013) Designing and benchmarking the MULTICOM protein structure prediction system. BMC Struct Biol 13:2
19. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P (2000) The protein data bank. Nucleic Acids Res 28(1):235–242
20. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402
21. Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. J mol Biol 215(3):403–410
22. Biegert A, Söding J (2009) Sequence context-specific profiles for homology searching. Proc Natl Acad Sci USA 106(10):3770–3775
23. Hughey R, Krogh A (1995) SAM: sequence alignment and modeling software system. In: Technical Report: UCSC-CRL-95-07. University of California at Santa Cruz
24. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39(Suppl 2):W29–W37
25. Soding J, Biegert A, Lupas A (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33(Web Server Issue):W244–W248
26. PRC, the profile comparer. http://supfam.org/PRC/
27. Sadreyev R, Grishin N (2003) COMPASS: a tool for comparison of multiple protein alignments with assessment of statistical significance. J Mol Biol 326(1):317–336
28. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33(7):2302–2309
29. Zhou H, Zhou Y (2005) SPEM: improving multiple sequence alignment with sequence profiles and predicted secondary structures. Bioinformatics 21(18):3615–3621
30. Fiser A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol 374:461–491

31. Cheng J, Wang Z, Eickholt J, Deng X (2011) Recursive protein modeling: a divide and conquer strategy for protein structure prediction and its case study in CASP9. In: Bioinformatics and Biomedicine Workshops (BIBMW). IEEE. p 352–357
32. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol 487:545–574
33. Wang Z, Tegge AN, Cheng J (2009) Evaluating the absolute quality of a single protein model using structural features and support vector machines. Proteins 75(3):638–647
34. Cheng J, Randall A, Sweredoski M, Baldi P (2005) SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res 33(Web Server Issue):W72–W76
35. Tegge AN, Wang Z, Eickholt J, Cheng J (2009) NNcon: improved protein contact map prediction using 2D-recursive neural networks. Nucleic Acids Res 37(Suppl 2):W515–W518
36. Cheng J, Baldi P (2005) Three-stage prediction of protein β-sheets by neural networks, alignments and graph algorithms. Bioinformatics 21(Suppl 1):i75–i84