# Integration of Template-Based and Template-Free Model Sampling for Protein Tertiary Structure Prediction by MULTICOM-NOVEL Server

Jilong Li, Badri Adhikari, and Jianlin Cheng*

*Department of Computer Science, University of Missouri, Columbia, MO 65211, USA*

chengji@missouri.edu

Our tertiary structure prediction server MULTICOM-NOVEL participated in the CASP11 experiment. The server used a novel integration of a variety of template-based and template-free model sampling methods.

## Method

MULTICOM-NOVEL generated an ensemble of protein models for each target using multiple different methods in each step of protein tertiary structure prediction, such as sequence/profile comparison tools (e.g., PSI-BLAST[1], HHSearch[2], RaptorX[3]), target-template alignment tools (e.g., PSI-BLAST[1], HHSearch[2], RaptorX[3], MSACompro[4], HHMsato[5], Promals3d[6]), MULTICOM template and alignment combination protocol[7,8], and model generation tools (e.g., our in house template-based MTMG, Modeller[9], local RaptorX[3], ab initio Rosetta[10], local I-TASSER[11,12], our in house *ab initio* contact-based CNS[13,14]). The ensemble of models of a target was evaluated by two methods: the single-model absolute model quality assessment tool – ModelEvaluator[15] and the fully pairwise model comparison tool – APOLLO[16]. From the ensemble, MULTICOM-NOVEL chose top five models ranked by the weighted sum of APOLLO scores and ModelEvaluator scores as final predictions.

In comparison with our methods tested in CASP10, the major new developments in MULTICOM-NOVEL fully or partially benchmarked in CASP11 include: 1) a novel probabilistic multi-template based model generation tool (MTMG); 2) a practical multiple protein sequence alignment algorithm MSACompro[4] using predicted secondary structure, solvent accessibility, and residue-residue contacts; 3) a novel and practical profile-profile pairwise protein sequence alignment algorithm HHMsato[5] based on hidden Markov models, secondary structure, solvent accessibility, torsion angle and evolutionary coupling information; 4) a new template-free modelling tool CNS using residue-residue contacts predicted by NNcon[17] and SVMcon[18] and secondary structure predicted using PSPro[19] as restraints with distance geometry simulated annealing protocol implemented in CNS software suite[13,14]. These new tools improve the quality and diversity of models generated in the ensemble.

## Results

We preliminarily evaluated MULTICOM-NOVEL on the whole chains of 36 CASP11 targets whose experimental structures were released to date. **Table 1** reports the average GDT-TS scores and TM-scores of top 1 and best of 5 models predicted by MULTICOM-NOVEL.

**Table 1**. The average GDT-TS scores and TM-scores of top one and best of five models on 36 CASP11 targets. These targets are T0759, T0760, T0761, T0762, T0763, T0764, T0765, T0766, T0767, T0768, T0769, T0770, T0771, T0772, T0774, T0776, T0780, T0781, T0782, T0784, T0786, T0790, T0791, T0800, T0801, T0808, T0815, T0831, T0833, T0834, T0845, T0849, T0852, T0853, T0855, and T0857.

| Predictor | Top One | | | Best of Five | |
|---|---|---|---|---|---|
| | GDT-TS | TM-score | | GDT-TS | TM-score |
| MULTICOM-NOVEL | 0.47 | 0.54 | | 0.50 | 0.58 |


1. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25, 3389-3402 (1997).
2. Soding, J., Biegert, A. & Lupas, A. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research 33, W244-W248 (2005).
3. Källberg, M. et al. Template-based protein structure modeling using the RaptorX web server. Nature protocols 7, 1511-1522 (2012).
4. Deng, X. & Cheng, J. MSACompro: Protein Multiple Sequence Alignment Using Predicted Secondary Structure, Solvent Accessibility, and Residue-Residue Contacts. BMC Bioinformatics 12, 472 (2011).
5. Deng, X. & Cheng, J. Enhancing HMM-based protein profile-profile alignment with structural features and evolutionary coupling information. BMC bioinformatics 15, 252 (2014).
6. Pei, J., Kim, B.-H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic acids research 36, 2295-2300 (2008).
7. Wang, Z., Eickholt, J. & Cheng, J. MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8. Bioinformatics 26, 882-888 (2010).
8. Li, J., Deng, X., Eickholt, J. & Cheng, J. Designing and Benchmarking the MULTICOM Protein Structure Prediction System. BMC Structural Biology, in press (2013).
9. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology 234, 779-815 (1993).
10. Leaver-Fay, A. et al. ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol 487, 545-574 (2011).
11. Zhang, Y. I-TASSER server for protein 3D structure prediction. BMC bioinformatics 9, 40 (2008).
12. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols 5, 725-738 (2010).
13. Brunger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallographica Section D: Biological Crystallography 54, 905-921 (1998).
14. Adhikari, B., Bhattacharya, D., Deng, X., Li, J. & Cheng, J. in The workshop on artificial intelligence and robotics methods in computational biology of 27th AAAI Conference, Bellevue, WA, USA.

15. Wang, Z., Tegge, A. N. & Cheng, J. Evaluating the absolute quality of a single protein model using structural features and support vector machines. Proteins: Structure, Function, and Bioinformatics 75, 638-647 (2009).
16. Wang, Z., Eickholt, J. & Cheng, J. APOLLO: a quality assessment service for single and multiple protein models. Bioinformatics 27, 1715-1716 (2011).
17. Tegge, A. N., Wang, Z., Eickholt, J. & Cheng, J. NNcon: improved protein contact map prediction using 2D-recursive neural networks. Nucleic acids research 37, W515-W518 (2009).
18. Cheng, J. & Baldi, P. Improved residue contact prediction using support vector machines and a large feature set. BMC bioinformatics 8, 113 (2007).
19. Cheng, J., Randall, A., Sweredoski, M. & Baldi, P. SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Research 33, W72-W76 (2005).