TopModel: A multiple-template meta-approach to homology modeling

Daniel Mulnaes, Holger Gohlke

Institute of Pharmeceutical and Medicinal Chemistry,
Department of Mathematics and Natural Sciences,
Heinrich-Heine University, Duesseldorf

Protein structure prediction is one of the most important problems in computational biology, and is key to understanding structural evolution, protein function, protein-ligand and protein-protein interactions, and for data-driven protein- or drug-design [1,2]. However, the current rate of structure determination by experimental methods is far exceeded by that of next-generation sequencing, to the point where only for 1/1000th of the known protein sequences a structure is known. For this problem, computational structure prediction is considered the best solution [3]. In the last decade, many methods have been developed in each of the computational fields necessary for automated structure prediction. The bi-annual community-wide blind experiment CASP evaluates the accuracy of work flows that integrate a subset of these methods, and has shown the value of meta- and consensus-approaches [3,4]. This work presents TopModel, a multiple-template meta-approach to homology modeling, which combines multiple state-of-the-art threading, alignment, and model quality assessment programs to provide a versatile work flow and toolbox for structure prediction. TopModel yields high-quality structures [5,6] and performs well even for low sequence identities. When benchmarked against CASP10 targets, TopModel shows an accuracy above the average state-of-the-art work flow. We anticipate that as more programs are integrated, the accuracy and sensitivity of our methods will improve. Hence, we aim to expand TopModel to achieve more accurate modeling of proteins with multiple domains and protein-protein complexes as well as to use it to evaluate alignment and threading software in a consistent manner based on the quality of models produced by the generated alignments.

[1] Dalton J.A., Jackson R.M., *Bioinformatics,* **2007**, *23*, 1901-1908.
[2] Biasini M., ... Schwede T. et al., *Nucleic Acids Res.*, **2014**, Web-edition, 252-258.
[3] Moult J, ... Tramontano A. et al., *Proteins,* **2014**, 82 Suppl 2, 1-6
[4] Rychlewski L., Fischer D., *Protein Science*, **2005**, 14, 240-245.
[5] Widderich N., ... Mulnaes D., ... Gohlke H., ... Bremer E, *J. Mol. Biol.* **2014**, 426, 586-600
[6] Gohlke H., ... Mulnaes D., ... Schmitt L., *J. Chem. Inf. Model.* **2013**, 53, 2493-2498