MS30.O02

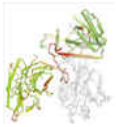
Expanding our knowledge of the protein universe: Modelling of protein structures

J. Haas[1,2], A. Barbato[1,2], T. Schmidt[1,2], S. Roth[1,2], A. Waterhouse[1,2], S. Bienert[1,2], K. Arnold[1,2], L. Bordoli[1,2], T. Schwede[1,2]

[1]University of Basel, Biozentrum, Basel, Switzerland, [2]SIB Swiss Institute of Bioinformatics, Computational Structural Biology, Basel, Switzerland

Computational modeling and prediction of three-dimensional macromolecular structures and complexes from their sequence has been a long standing goal in structural biology. Over the last two decades, a paradigm shift has occurred: starting from a large "knowledge gap" between the huge number of protein sequences compared to a small number of experimentally known structures, today, some form of structural information – either experimental or computational – is available for the majority of amino acids encoded by common model organism genomes. Methods for structure modeling and prediction have made substantial progress of the last decades, and template based homology modeling techniques have matured to a point where they are now routinely used to complement experimental techniques. However, computational modeling and prediction techniques often fall short in accuracy compared to high-resolution experimental structures, and it is often difficult to convey the expected accuracy and structural variability of a specific model. Retrospectively assessing the quality of blind structure prediction in comparison to experimental reference structures allows benchmarking the state-of-the-art in structure prediction and identifying areas which need further development. The Critical Assessment of Structure Prediction (CASP) experiment has for the last 20 years assessed the progress in the field of protein structure modeling based on predictions for ca. 100 blind prediction targets per experiment which are carefully evaluated by human experts. The "Continuous Model EvaluatiOn" (CAMEO) project aims to provide a fully automated blind assessment for prediction servers based on weekly pre-released sequences of the Protein Data Bank PDB. CAMEO has been made possible by the development of novel scoring methods such as lDDT, which are robust against domain movements to allow for automated continuous structure comparison without human intervention.

[1] *T. Schwede, Structure, 2013, 21, 1531-1540.,* ***[2]*** *V. Mariani, M. Biasini, A. Barbato, et al., Bioinformatics, 2013, 29, 2722-2728.,* ***[3]*** *J. Moult, K. Fidelis, A. Kryshtafovych, et al., Proteins, 2014, 82(S2), 1-6.*



Keywords: Modelling, Protein structure prediction, Bioinformatics