Residue-residue contacts in modeling protein structure

M. Kotulska[1]*, P.P. Wozniak[1], B. Konopka[1], M. Kurczynska[1], G. Vriend[2],
A. Sokolowska[1], J. Wojciechowski[1]

[1] *Wroclaw University of Science and Technology, Faculty of Fundamental Problems of Technology, Department of Biomedical Engineering, Wroclaw, Poland*
[2] *Radboud University Medical Centre, Centre for Molecular and Biomolecular Informatics, Nijmegen, The Netherlands*
* e-mail: malgorzata.kotulska@pwr.edu.pl

Key words: correlated mutations, contact maps, mirror chirality

Motivation and Aim: Experimental study of the molecular structure is expensive, time consuming and not always feasible. The difficulty with experiments results in a great disproportion between the numbers of known structures and protein sequences. Bioinformatical modeling of a protein predicts its molecular structure. Various computational methods use different approaches and generate different structural models of the same protein. Also models from the same method can differ. Choosing the best model is a difficult task, especially that accuracy of each method is determined based on its average efficiency related to available experimental structures. It does not necessarily correspond to the situation of modeling an individual unknown protein. Information of contact distribution supports the modeling as it brings additional different information.

Methods and Algorithms: We applied Direct Contact Analysis (DCA) methods, which use correlated mutations of amino acids. Obtained contact maps were first used to support prediction of molecular structure. The studies regarded characteristics of the results and obstacles appearing in modeling of protein structures based on full and incomplete contact maps. The maps resulting from modeling are incomplete mainly because of insufficient accuracy of current methods. We proposed several methods to improve it, based on machine learning and structural filters, as well as a method to forecast efficiency of contacts prediction for an individual protein.

Results: The studies showed that for correct protein reconstruction 30 % of random contacts is sufficient, however it depends on the protein class. Density of contacts and their positions relative to the interior of the molecule are also important in proper reconstruction [1]. Contact maps do not provide all the necessary information though, allowing for models of mirror chirality difficult to filter out. The knowledge of even incomplete maps allows selecting misfolded from properly folded proteins [2]. We designed a regression model that forecasts the accuracy of contact prediction for individual proteins with an average error of 7 percentage points [3]. Contact prediction can be further raised by including machine learning methods and structural filters.

Conclusion: The results obtained in our studies show potential of contact maps in modeling protein structures.

References
1. Konopka B.M. et al. (2014) Automated procedure for contact-map-based protein structure reconstruction. Journal Membrane Biol. 247(5):409-420.
2. Wozniak P.P. et al. (2017) Forecasting residue-residue contact prediction accuracy. Bioinformatics. 33(21):3405-3414.
3. Wozniak P.P. et al. (2017) Correlated mutations select misfolded from properly folded proteins. Bioinformatics. 33(10):1497-1504.