PeerJ Preprints NOT PEER-REVIEWED

# Correlated mutations distinguish misfolded and properly folded proteins

Pawel P. Wozniak[1], Malgorzata Kotulska[1*], Gert Vriend[2]

[1]Department of Biomedical Engineering, Wrocław University of Science and Technology, Wrocław, Poland
[2]Centre for Molecular and Biomolecular Informatics, Radboudumc, Nijmegen, the Netherlands

*Corresponding Author: malgorzata.kotulska@pwr.edu.pl

Knowledge about the three dimensional structure of proteins is crucial in order to learn about their behavior, stability, or role as a target in drug design. Unfortunately, traditional experimental methods used in structure determination such as X-ray crystallography and NMR are costly and time-consuming. Therefore, computational methods that allow for protein structure reconstruction from sequence only are greatly desired. One of these is the recently developed direct coupling analysis (DCA) method [1, 2] which achieves the best results in residue-residue contact prediction from multiple sequence alignments only. Predicted contacts are used as restraints in the reconstruction of the three-dimensional structure of a protein. Unfortunately, the accuracy of DCA methods is on the order of 40% among the 100 strongest predicted contacts, which is insufficient for ab initio protein structure reconstruction. However, the results of DCA can support protein structure reconstruction in a different way.

Our results show that DCA can indicate the best protein structure among its structural variants by the prediction of residue-residue contacts [3]. We counted the number of correctly predicted contacts within the strongest 100 DCA predictions for a set of obsolete PDB entries and their successors and for 22 proteins for which the Decoys 'R' Us database [4] provided properly folded and misfolded structures. These numbers were related to structure similarity scores, such as RMSD or TM-score [5]. DCA correctly predicts significantly more contacts for properly folded structures than for misfolded ones. Our method works much better for structures determined with X-ray crystallography than with the NMR spectroscopy [3]. The method will not detect misfolded proteins per se, but when a protein structure experimentalist needs to choose between alternative folds for the same protein, DCA can help.

[1] F. Morcos et al., Direct-coupling analysis of residue coevolution captures native contacts across many protein families, 2011, Proc Natl Acad Sci U S A 108(49):E1293-301.
[2] C. Feinauer et al., Improving contact prediction along three dimensions, 2014, PLoS Comput Biol., 10(10):e1003847.
[3] P.P. Wozniak, G. Vriend, M. Kotulska, Correlated mutations select misfolded from properly folded proteins, 2016, Bioinformatics, (article accepted for publication on 10th January 2017).
[4] R. Samudrala, M. Levitt, Decoys 'R' Us: A database of incorrect protein conformations to improve protein structure prediction, 2000, Protein Science 9: 1399-1401.
[5] Y. Zhang, J. Skolnick, TM-align: A protein structure alignment algorithm based on TM-score, 2005, Nucleic Acids Research, 33: 2302-2309.