Correlated mutations select misfolded from properly folded proteins

Wrocław University of Science and Technology

Pawel.P.Wozniak@pwr.wroc.pl[1], Gert Vriend[2], Małgorzata Kotulska[1]

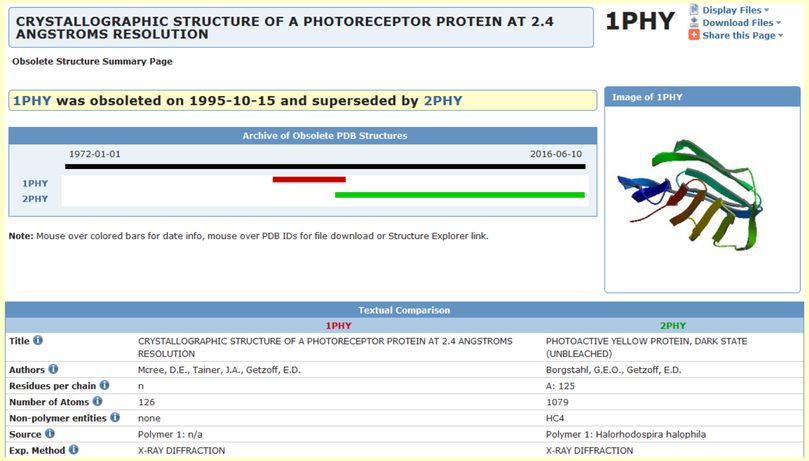
[1]Department of Biomedical Engineering, Wroclaw University of Science and Technology, Poland

[2]Centre for Molecular and Biomolecular Informatics, Radboud University Nijmegen Medical Centre, The Netherlands

Main objective:

Indication of a better structure among properly folded and misfolded variants of a protein with residue-residue contacts predicted with DCA algorithm.

Obsolete and successor structures



CRYSTALLOGRAPHIC STRUCTURE OF A PHOTORECEPTOR PROTEIN AT 2.4 ANGSTROMS RESOLUTION

1PHY Display Files Download Files Share this Page

Obsolete Structure Summary Page

1PHY was obsoleted on 1995-10-15 and superseded by 2PHY

Image of 1PHY

Archive of Obsolete PDB Structures

1972-01-01 2016-06-10

1PHY

2PHY

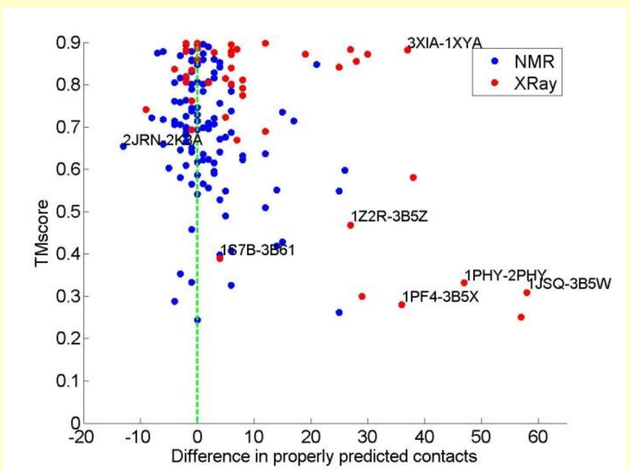
Note: Mouse over colored bars for date info, mouse over PDB IDs for file download or Structure Explorer link.

Textual Comparison

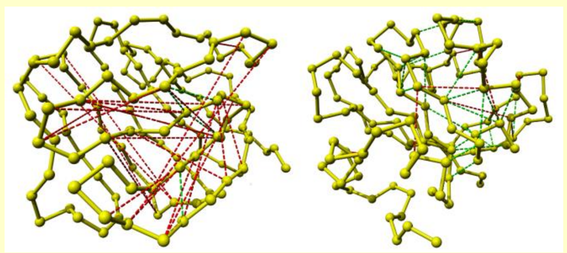
	1PHY	2PHY
Title	CRYSTALLOGRAPHIC STRUCTURE OF A PHOTORECEPTOR PROTEIN AT 2.4 ANGSTROMS RESOLUTION	PHOTOACTIVE YELLOW PROTEIN, DARK STATE (UNBLEACHED)
Authors	Mcree, D.E., Tainer, J.A., Getzoff, E.D.	Borgstahl, G.E.O., Getzoff, E.D.
Residues per chain	n	A: 125
Number of Atoms	126	1079
Non-polymer entities	none	HC4
Source	Polymer 1: n/a	Polymer 1: Halorhodospira halophila
Exp. Method	X-RAY DIFFRACTION	X-RAY DIFFRACTION

There are almost 2500 obsolete structures in the PDB [1]. Most of them were superseded and replaced by a successor structure. Usually a successor structure is a further refined version of the previous one and does not differ much from the superseded PDB entry. Occasionally, the obsolete structure was misfolded, and got replaced by a more correct version. One example is a photoactive yellow protein deposited in 1989 as a beta-clam (1PHY). After six years, the misinterpretation of electron density was discovered and structure became entry 2PHY, which is an alpha/beta-fold.

Individual results



(left figure) Several obsolete-successor pairs were investigated individually. Differentiating contacts were usually related to the important functional sites of the proteins. Positive difference in properly predicted contacts means more contacts properly predicted for the successor structure.

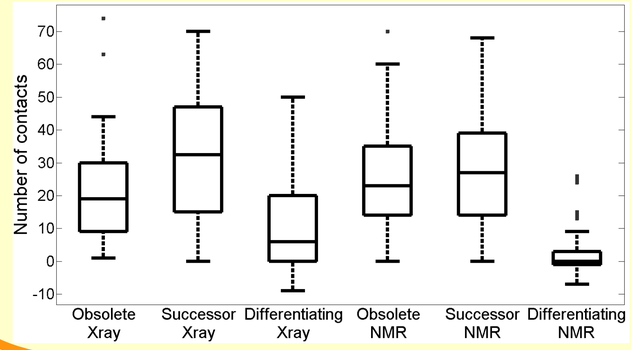


(right figure) Contacts predicted for 1PHY (left) and 2PHY (right) structures. Proper and wrong contacts were shown in green and red, respectively.

Methodology

- Two residues are regarded in contact when a distance between their Cα atoms is less than 8 Å and when they are separated in the sequence by at least 4 positions.
- The list of almost 2500 obsolete and successor pairs was obtained from the PDB.
- Sequences shorter than 40 amino acids and obsolete-successor pairs with the RMSD lower than 1.1 Å and a TM-score [2] above 0.9 were removed (these cut-offs were manually optimized). Pairs for which coordinates were not available in PDB format were also removed.
- The final dataset consisted of **152** pairs - **42** from X-ray and **110** from NMR. Only the first model was used from each NMR structure ensemble.
- gplmDCA [3] was used to calculate DCA scores for pairs of residues. Input MSAs for gplmDCA were created with HHblits.
- Pairs of positions for each protein were sorted in descending order of the DCA score. The 100 strongest predicted pairs of residue positions were examined.
- Predicted contacts that were observed as real contacts in either the proper structure or its misfolded alternative (but not in both) were called differentiating contacts, unless the difference between their contact distances was less than 1 Å.

Global results



Numbers of correctly predicted contacts for obsolete and successor structures and for the differentiating contacts in each pair. Positive number of differentiating contacts means more contacts properly predicted for the successor structure.

Conclusions

- DCA can be used successfully to discriminate between misfolded protein structures and their properly folded successor structures.
- Proper structure indication is more accurate for protein structures determined with X-ray crystallography than NMR spectroscopy. This is probably related to the fact that NMR structures are on average twice shorter than X-ray structures.
- Differentiating contacts not only indicate the proper structure but also are usually related to the functionally significant sites of proteins such as active sites or hydrophobic cores.

[1] Berman HM, Westbrook Z, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res., 28:235-242

[2] Zhang Y, Skolnick J (2005) TM-align: A protein structure alignment algorithm based on TM-score , Nucleic Acids Research, 33: 2302-2309.

[3] Feinauer C, Skwark MJ, Pagnani A, Aurell E (2014) Improving contact prediction along three dimensions. PLoS Comput Biol., 10(10):e1003847.