

Protein Designability and Disease

Institute of Bioinformatics

Biological structures often occur at different frequencies. Some structures are more widespread than others. Why are some structures more successful than others? We investigate this question with respect to protein domains.

Based on the designability hypothesis, robustness against mutation is a deciding factor of success of the structure. Mutations affecting proteins may involve substitutions, deletions, insertions and even duplications of sequences encoding and regulating the protein. Protein domains sharing overall structural similarity are termed „folds“. We estimate the designability of a particular fold by the number of families known to form that fold.

In accordance with our hypothesis, we found that folds predicted to be more designable are on average not only more widespread in human, mouse and yeast but have proliferated at a greater rate since the time of the common ancestor of these organisms.

Current research also suggests that most, if not all, known hereditary disease causing mutations in coding regions affect protein structure. Because more designable structures are expected to be more robust against mutations, we investigated whether hereditary disease-related proteins were predicted to be less designable.

We find that relatively large numbers of proteins annotated with OMIM (Online Mendelian Inhe-



Philip Wong

Institute of Bioinformatics

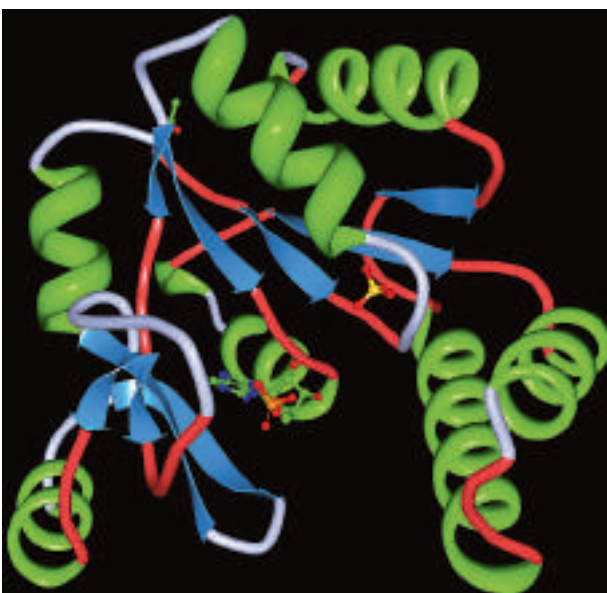
Telephone +49 - 89 / 31 87-35 77
philip.wong@gsf.de

ritance in Man) associated diseases have folds with low numbers of families. In particular, protein structures of a lesser predicted designability were found to be associated with very frequent diseases.

Our results suggest that designability is an important determinant of the distribution of structures in nature and their relation to human disease.

Literature:

- Wong P, Frishman D.: Designability, Distribution, and Disease. PLoS Comput. Biol. 2, e40, 2006.



Example of a protein with the highly ordered c.37 fold (P-loop with nucleoside hydrolase) with 22 known families.