The coverage of protein fold recognition

Walter A. Koppensteiner, Markus Wiederstein,
Peter Lackner and Manfred J. Sippl

Center of Applied Molecular Engineering, University of Salzburg
Jakob-Haringer-Straße 3, A-5020 Salzburg, Austria
walter@came.sbg.ac.at

Background: The protein fold recognition technique identifies for a given sequence a native-like template in the pool of existing structures. Consequently, it can only predict proteins whose structures are similar to existing ones. The minimal requirement for a structure to be a suitable template for modeling by fold recognition is to share at least all core secondary structure elements with the native structure. To assess the applicability of fold recognition it is necessary to determine the fraction of proteins which fulfill this requirement.

Results: We have analyzed the structure relationships of all proteins of the PDB which were released in 1998 and did not have a significant sequence similarity to an existing structure at the time of release. The 174 proteins (196 domains) of this set where superimposed to a representative subset of the PDB using the structure comparison program ProSup [1]. When super-imposing the alpha carbon traces it turned out that 76% of the domains have significant similarity to an existing structure, 11% can be regarded as novel folds while the rest share some similarity to an existing structure. The picture changes when also the positions of beta carbon atoms are taken into consideration for structure superimposition. Then only 71% have significant similarity to existing structures but 19% are novel folds. The inclusion of beta carbon atoms makes structure comparison more restrictive because pairs with opposite side chain orientations are no longer considered as equivalent. However, side chain orientation is an important issue to derive an accurate model of a protein structure. Therefore, structure comparison with beta carbon atoms is a reasonable means to measure the fraction of recognizable templates. From this study, we conclude that at present 71% of the proteins whose structure could be determined experimentally share enough similarity with an existing domain that the latter is a suitable template for fold recognition.

Reference:

[1] Feng, Z.-K. and Sippl, M.J. (1996) Optimum super-imposition of protein structures: ambiguities and implications. *Folding & Design*, 1(2): 123-132