Poster I-64

Structural classification and prediction of reentrant regions in alpha-helical transmembrane proteins: application to complete genomes

Authors:
Håkan Viklund (*Stockholm University*)
Erik Granseth (*Stockholm University*)
Arne Elofsson (*Stockholm University*)

Short Abstract: 3D-structures of transmembrane proteins have been used to define and study the concept of reentrant regions. The results include statistics for the amino acid composition of reentrant regions, a novel prediction method and estimations of the occurence of reentrant regions in three genomes.

Long Abstract:
All living cells are abundant in alpha-helical transmembrane (TM) proteins. According to earlier estimates they make up about 20-30% of a typical genome and are vital for such processes as endocytosis and exocytosis, intracellular transport, cell-cell interaction mechanisms of multicellular organisms and receptor mediated signaling.

To better understand these mechanisms, further insight into the structural properties of alpha-helical TM proteins is important. Over the last few years, there have been a number of studies examining the properties mainly of the membrane spanning alpha-helices, describing such things as their general amino acid composition, the presence of specific sequence patterns and motifs, the packing of helix bundles and the properties of inter-helical residue interactions. Recently, there have also been studies focusing on some of the structural properties of the loop regions, among other things the parts of the proteins located at the membrane-water interface region.

In this study we continue the examination of the structural features of loop regions by performing the first overall analysis of the properties of membrane penetrating regions that are not TM helices, referred to as reentrant regions. This type of structural region has been observed in a number of different TM proteins. A typical example is the ``half-TM'' reentrant loops of aquaporin like structures, which have important functional roles both in preventing proton conductivity and as selectivity filters.

Using a dataset of 79 chains from the Protein Data Bank, where no one chain is more than 30% identical to any other in the set, we define the concept of reentrant regions using purely structural criteria. This results in a set of 36 reentrant regions containing a considerable variance both in length, penetration depth and overall shape, but with a surprisingly high homogeneity regarding the amino acid composition. In addition, we have constructed TOP-MOD, a simple HMM-based method that focuses on predicting reentrant regions, and used it to investigate the occurrence of reentrant regions in the genomes of Escherichia coli, Saccharomyces cerevisiae and Homo sapiens.

With reservation for the limited amount of data, we find that reentrant regions can be divided into three distinct categories based on their secondary structure content and that they can be detected in a TM protein sequence with reasonably high accuracy based on their amino acid composition, the strongest signal being that the residues in reentrant regions are considerably smaller on average compared to other regions. Further, we find that the fraction of proteins containing reentrant regions seems to be >10% in a typical genome and that it increases linearly with the number of TM regions. In agreement with earlier observations, we find that reentrant regions seem to be most commonly found in channel proteins and least commonly in receptors, but contrary to earlier observations they seem to be fairly common also in transporter proteins.