A structural study for the optimization of functional motifs encoded in protein sequences - (session: Structural Genomics)

A. Via, M. Helmer-Citterich

Centre for Molecular Bioinformatics, Dept. of Biology, University of Rome Tor Vergata

Many PROSITE sequence motifs match all and only the known true positive sequences. However a great number of PROSITE motifs (referred in the following as „leaky‰ patterns) select false positives and/or miss known true positives. It is possible that ˆ at least in some cases - the weak specificity and/or sensitivity of a pattern is due to the lack of key residues in the sequence pattern. Indeed, while localized in space, functional residues may be dispersed along the protein sequences and therefore not easily detectable in multiple sequence alignments. In the present work, a set of eight PROSITE deterministic patterns was selected among the „leaky‰ ones and the structures of the corresponding true positives were analyzed by means of the 3D profile procedure (de Rinaldis et al., 1998) and by visual inspection. For each considered pattern, this approach allowed the identification of sets of residues structurally well conserved in the protein surface region nearby the PROSITE residues. In seven out of eight cases only a subset of the conserved amino acids belongs to the original PROSITE pattern while the remaining part falls outside. Only in one case (CYTOCHROME_C) all the structurally conserved residues belong to the original pattern. The structurally conserved residues falling outside the PROSITE patterns were used to build what we called EXTENDED sequence patterns. The CYTOCHROME_C pattern was discarded.
RESULTS: EXTENDED patterns were matched against the SWISS-PROT database to test their sensitivity and specificity. In all the analyzed cases, the addition of information inferred from structural analysis improved pattern specificity and in some cases both specificity and sensitivity.