Derivation and refinement of global sequence motifs for the integral membrane proteins

Michael Sadowski, J. Howard Parish and David R. Westhead

Previously, an alignment algorithm for matching flexible, sequence-length motifs ('signatures') to protein sequences has been described and a method of deriving these motifs using contact information derived from the crystal structures of various protein families was reported subsequently.

This project extends the previous work by providing a method of deriving such motifs for families where little or no structural information is available. Multiple sequence alignments constructed using predicted secondary structure information are created for a minimal set of sequences from the family of interest, motifs being selected from these on the basis of conservation within columns of the alignment.

At present, work has focused on the G-protein coupled receptors, a superfamily of proteins responsible for a large amount of hormonal and neural signalling in eukaryotes. Using the above method motifs have been derived which can identify up to 75% of superfamily members before the first false positive sequence when aligned to the SWISS-PROT 39 database. An adaptive method based on the Genetic Algorithm of Holland (1975) has been developed for the refinement of these motifs which is successful but requires a great deal of time to produce results.

In future, the project will be extended in the following ways:

1) Other families of membrane proteins will be examined using the above method for selection of motifs from multiple alignments.
2) Derivation of signatures using the Genetic Algorithm method will be attempted using a small subset of the SWISS-PROT database.
3) A number of strategies for reducing the time required for refinement of signatures will be investigated, including the development of reduced versions of the SWISS-PROT database.
4) The statistics of signature scoring will be investigated in order to improve the power of this technique for identifying protein families.

References

Daniel, S.C., Parish, J.H., Ison, J.C. & Findlay, J.B.C. (1999) Alignment of a sparse protein signature with protein sequences: application to fold recognition for three small globulins. *FEBS Lett.* **459**, 349-352

Ison, J.C., Blades, M.J., Bleasby, A.J., Daniel, S.C., Parish, J.H. & Findlay, J.B.C. (2000): Key residues approach to the definition of protein families and analysis of sparse family signatures. *Proteins* **40**, 330-341

Funding

We thank the MRC for funds to support this work.