InterPro: Integrated Resource Of Protein Domains and Functional Sites

Alexander Kanapin, Wolfgang Fleischmann, Evgueni Zdobnov, Rolf Apweiler.

EMBL Outstation - The European Bioinformatics Institute,
Hinxton, Cambridge CB10 1SD, UK.

alexander.kanapin@ebi.ac.uk

Keywords: PROSITE, Pfam, PRINTS, ProDom, SWISS-PROT, TrEMBL.

The databases PROSITE, Pfam, PRINTS, ProDom, SWISS-PROT and TrEMBL joined forces to launch a new Integrated Resource of Protein Domains and Functional Sites, abbreviated InterPro.

While the participating databases will maintain their indiviual approaches with all the known benefits, InterPro will serve as a common co-ordinate system, harmonizing domain definitions, nomenclature,annotation, match-lists and hyperlinks. This collaboration allows us to provide a convenient one-stop search engine to scan your protein sequencesagainst all known profiles, hidden Markov models, patterns andfingerprints. Up to now, it was difficult to compare hits to the different databases,as they are based on different protein database versions. This synchronisation problem has been solved. The above mentioned motif databaseswill continue with their release schedules, while the time between releases is covered by the EBI on a weekly basis.

InterPro entries contain links to the motif databases, a general description, method specific descriptions, references, and a list of matched proteins. Every entry is classified as describing a protein family, a domain, or a post-translational modification site. Furthermore, we maintain a list of parent-child relationships and link entries which match protein families and subfamilies or domains and more specific sub-domains.

To provide early access to InterPro, the available beta release contains manually checked links and concatenated annotation taken from the participating databases. In due course, the annotation will be combined and rewritten by the InterPro consortium, aiming at a reference source of information on protein sites, domains, and families.

The unified search engine, the relational schema, and the documentation on InterPro are available at www.ebi.ac.uk/interpro, the flatfile at ftp.ebi.ac.uk/pub/databases/interpro, and the sequence retrival system (SRS) version at srs.ebi.ac.uk.