

Poster B-37

InterProScan New and Planned Features



Authors:

Sarah Hunter (*European Bioinformatics Institute*)
Nicola Mulder (*European Bioinformatics Institute*)
Alberto Labarga (*European Bioinformatics Institute*)

Short Abstract: InterProScan is arguably the best-known tool for predicting protein functional annotation. It is a perl-based software package which wraps search algorithms for 10 protein signature databases and 4 other programs. This poster describes the architecture of the software, together with new and up-and-coming features.

Long Abstract:

Protein signatures are useful tools for classifying proteins into families and identifying domains. Ten of the major protein signature databases are integrated into InterPro, a database which provides such a classification for all Uniprot proteins. Each protein signature database has independently developed search algorithms for running proteins against their models in order to find those that match. They typically use standard software, but with additional pre- and post-processing scripts unique to each database.

In order to generate this information for the InterPro database, and rather than having to search each member database separately, a perl-based software package known as InterProScan was developed. It is now widely used in the scientific community for functional annotation of proteins and genomes. InterProScan wraps the search algorithms for each of the signature databases into a single package that allows a user to seamlessly search the databases integrated within InterPro with protein or nucleotide sequence(s). Users can do this via the command-line or a web interface, receiving their final predicted results in a variety of formats, such as XML, HTML, plain text and GFF (General Feature Format). From the results it is easy to directly link to the extensive curated knowledge present in InterPro, including the popular InterPro-to-GO mappings, which allow for automatic annotation of proteins to GO terms.

InterProScan can be accessed in three different ways:

- Users can search with their sequence via the EBI website (<http://www.ebi.ac.uk/InterProScan/>)
- They can use the recently-developed Web Services API (<http://www.ebi.ac.uk/Tools/webservices/WSInterProScan.html>)
- For bulk or confidential searches, users can download and install the entire package to their local servers (<ftp://ftp.ebi.ac.uk/pub/software/unix/iprscan/index.html>).

The poster will describe the general architecture of the software, together with notable features and recent or planned changes for 2006. These changes include a new XML file

format, allowing more extensive information to be stored and displayed, plus greater flexibility when converting between formats. We will also describe the Web Services API developed for programmatic access to the tool. Finally, we will describe the current strategies for speeding up InterProScan both on the EBI website and for users who download the software. These strategies include the creation of a "slim" version of InterProScan which will only search a subset of the models but will still deliver the same high quality results.