Recent developments in UniProt: improving access to protein knowledge

What is UniProt?
Comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

What’s new in UniProt?

Redesigned website
New homepage and advanced search functionality
New results page interface with easy to use filters
Search results can be customised and shared
Basket to store favourite proteins and build up your own set
New protein entry page content classification and navigation bar
New tool output interfaces

Improved evidence attribution
Information in a UniProtKB record comes from a range of different sources. The origin of each piece of information in an entry is provided so that users can clearly see these data sources. UniProt makes use of a subset of evidence codes from the Evidence Code Ontology to indicate data origin. These ECO codes are shown directly in the text version of the entries. On the UniProt website, they are transformed into user-friendly, easy to understand evidences.

Annotation scores
Annotation scores provide a measure of the annotation content of a UniProtKB entry or proteome. A 5-point heuristic scoring system is used where a score of 5 is associated with the best-annotated entries. Annotation scores can help to give a quick idea of the relative level of annotation of the entries in a set of search results.

Reduced proteome redundancy
UniProtKB has grown exponentially in recent years due to the vastly increased submission of multiple genomes for the same or closely related organisms. To reduce this sequence redundancy, we have developed a procedure to identify redundant proteomes within species groups and redundant bacterial proteomes have been removed from UniProtKB. All proteomes remain searchable through the Proteomes interface and redundant proteome sets are available for download from UniParc.

New proteomics mapping files
Mappings of UniProtKB human sequences to identified human peptides from public mass spectrometry proteomics repositories can now be found in a new dedicated ‘proteomics_mapping’ directory on the UniProt FTP site. The human reference proteome contains 89383 sequences and our analysis provides mass spectrometry evidence for 68229 of these sequences.

Keep up with changes in UniProt
Read the UniProt blog
http://insideuniprot.blogspot.co.uk/
Follow UniProt on Twitter
@uniprot
Visit the UniProt Facebook page
Subscribe to the UniProt RSS feed

UniProt is funded by the European Molecular Biology Laboratory, National Institutes of Health, European Union, Swiss Federal Government, British Heart Foundation and National Science Foundation.

Email: help@uniprot.org
URL: www.uniprot.org