

Need help?

Support: goa@ebi.ac.uk

UniProtKB-GOA project
EMBL-EBI
Wellcome Trust Genome Campus
Hinxton, Cambridge
CB10 1SD, UK

Further reading

Barrell, D. *et al.* (2009) The GOA database in 2009 – an integrated Gene Ontology Annotation resource. *Nucleic Acids Res.* 37, D396–D403

The Gene Ontology Consortium. (2008) The Gene Ontology project in 2008. *Nucleic Acids Res.* 36, D440–D444

Bairoch, A. *et al.* (2005) The Universal Protein resource (UniProt). *Nucleic Acids Res.* 33, D154–D159

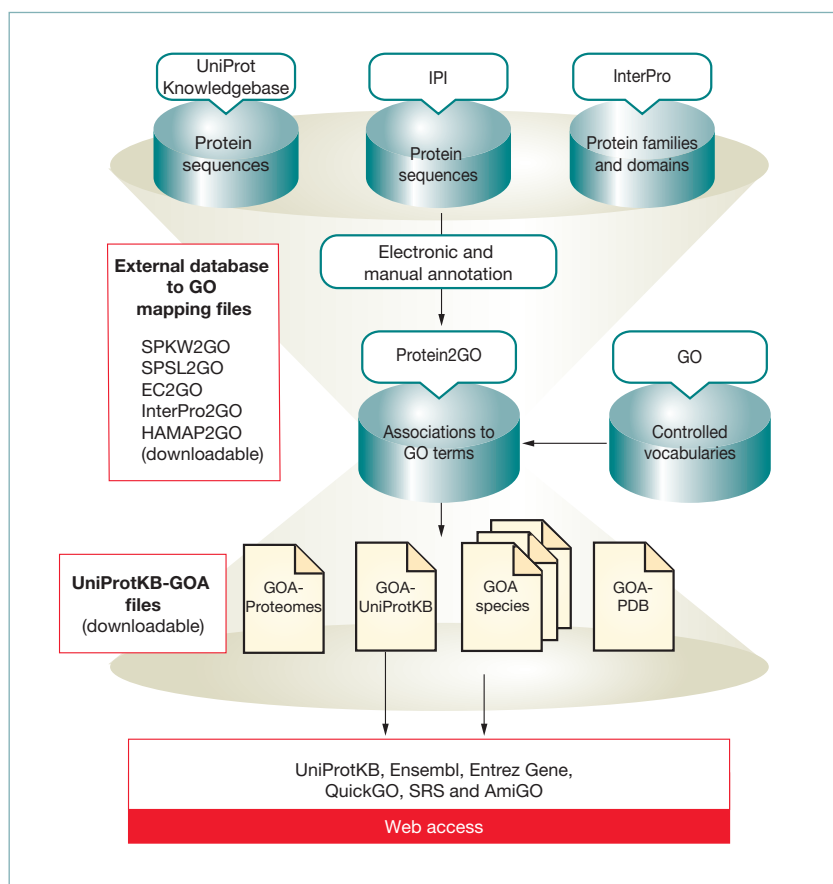
Gene Ontology Annotation project

www.ebi.ac.uk/GOA

As proteomics research gains momentum, biologists need new ways to access and analyse information on proteins. To exploit the potential of these data fully, we need to capture all the available biological information related to each protein, including consistent descriptions of protein function. The UniProtKB-Gene Ontology Annotation (UniProtKB-GOA) project uses the Gene Ontology (GO) – a dynamic controlled vocabulary that details the biological processes, molecular functions and cellular components of a generic cell – to describe proteins in the EBI's core genome and proteome databases (UniProtKB and the International Protein Index, IPI).

What is the UniProtKB-Gene Ontology Annotation project?

The UniProtKB-GOA project has assigned GO terms to all complete and incomplete proteomes that exist in UniProtKB, using a combination of electronic methods and manual curation. UniProtKB-GOA is updated on a monthly basis, in accordance with the latest data released by UniProtKB, IPI, Ensembl and InterPro (a documentation resource of protein families, domains and sites). By annotating all characterised proteins with GO terms and facilitating the transfer of this knowledge to similar uncharacterised proteins, we hope to contribute to a better understanding of all proteomes. The UniProtKB-GOA group is the GO Consortium member responsible for providing annotations to the human, chicken and bovine proteomes.



Sources and flow of data in UniProtKB-GOA.



How is GO annotated at the EBI?

The large-scale assignment of GO terms to UniProtKB entries involves a number of electronic techniques. We use existing information within database entries, including Swiss-Prot keywords (SPKW2GO) and cross-references to InterPro (InterPro2GO), which are manually mapped to GO terms. Electronically combining these mappings with matching UniProtKB entries generates a set of GO annotations. Secondly, in collaboration with the Ensembl team, manual GO annotations are projected automatically to a number of predicted orthologs using the Compara database.

High-quality manual annotations are made by highly-trained biological curators. UniProtKB-GOA also integrates manual annotations from all GO Consortium annotation groups, as well as a number of special interest databases (e.g. LIFEdb, Human Protein Atlas, Reactome and the IntAct database). For each GO association, an evidence code is included to categorise the type of information that has been found to support the annotation, for instance, associations that are made electronically are labelled as 'Inferred from Electronic Annotation'.

Retrieving data from UniProtKB-GOA

The UniProtKB-GOA project data can be accessed using several web-based browsers. The UniProtKB-GOA files can also be downloaded. For online links to the resources in the side bar (right), please go to www.ebi.ac.uk/GOA

What can I do with UniProtKB-GOA?

The success of GO can be measured by the number of databases that use it to annotate and exchange biological knowledge (see GO information sheet). The UniProtKB-GOA project has made an important contribution to this global effort. UniProtKB-GOA allows you to:

- access functional information for proteins in UniProtKB or Ensembl, either on an individual protein basis by querying the QuickGO or Ensembl browsers, or by downloading and parsing one of our gene association files;
- use GO Slims to summarise the biological attributes of a proteome, compare proteomes or, for example, find out what proportion of a proteome has been found to be involved in apoptosis;
- incorporate GO annotations into your own database to enhance the functional information available to your user community;
- use GO annotations to link between biological knowledge and high-throughput genomic or proteomic datasets;
- generate automated GO annotations to new genomic or protein sequences, using the InterProScan tool;
- find the location of human genes mapped to a particular GO term using Ensembl GOView.

For more information about applications of GO annotation please see the GO factsheet and studies cited at www.geneontology.org/cgi-bin/biblio.cgi

Contributing to the UniProtKB-GOA project

The success and accuracy of UniProtKB-GOA relies on frequent electronic and manual checking. If you find anything in UniProtKB-GOA that needs correcting or updating, please let us know by email, goa@ebi.ac.uk. Please provide the accession number of the protein entry in question, along with the source and annotation retrieval date. ●

UniProtKB-GOA resources

Web-based tools:

QuickGO

UniProtKB-GOA browser for the GO and GO annotations

www.ebi.ac.uk/QuickGO

AmiGO

Official GO Consortium browser for the GO and GO annotations

<http://amigo.geneontology.org/cgi-bin/amigo/go.cgi>

Ensembl www.ensembl.org

SRS <http://srs.ebi.ac.uk>

IPI www.ebi.ac.uk/IPI

Downloads:

UniProtKB-GOA gene association files

- GOA-UniProt, GO annotations for UniProtKB accessions for all species <ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/UNIPROT>

- Non-redundant GO annotations for human, mouse, rat, Arabidopsis, chicken, cow and zebrafish proteome sets:

<ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/...>

HUMAN
MOUSE
RAT
ARABIDOPSIS
CHICKEN
COW
ZEBRAFISH

Please note that two compressed files are provided in each of the above folders – a file of GO annotations and a file of cross-references, allowing users to map between different sequence identifiers.

- Species-specific collections of GO annotations to UniProtKB accessions for complete proteomes where the genome has been completely sequenced and is in the public domain

<ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/proteomes>

About GO

UniProtKB-GOA is supported by the European Molecular Biology Laboratory, the US National Institutes of Health (NIH), the British Heart Foundation and Kidney Research UK.