Overview

The Drosophila protein annotation program focuses on the expert annotation of characterised D.melanogaster proteins. The current UniProt release (2016_03 – Mar-16, 2016) contains 550,740 reviewed entries, of which 3,278 entries are from D. melanogaster. Each entry contains information integrated from a range of sources such as the scientific literature, protein sequence analysis tools, other databases and automatic annotation systems to provide an overview of available protein knowledge with clear indications of where that information comes from.

Why the fly?

Although they are very distantly related, most of the fundamental biological mechanisms and genetic pathways that control development and survival are conserved between flies and humans.

Expert curation of a Drosophila entry

Sequence curation
Sequences from the same gene are merged and analysed. Sequence analysis utilises a range of tools that produce computer predictions which are integrated into the entry along with experimental data reported in cited references.

Tools
Tools such as ‘Blast’ and ‘Add to basket’ aid further analysis for the user.

Sequence feature viewer
Sites and regions of interest are described and their positions can be viewed graphically.

Literature curation
Relevant publications are read in detail and fully curated. Information on function, expression, interactions, subcellular location, sequences, disruption phenotype etc. are extracted and added to the entry.

The quality of predicted and experimental data is manually assessed before integration, so that each entry provides a comprehensive and critical review of the most current biological information for a protein.

Cross-references and Links
Point users to information from sources such as INSDC, Refseq, PRIDE, Ensembl, KEGG, Flybase etc.


Link to the Drosophila proteome in UniProt.

All D. melanogaster entries are cross-referenced with Flybase to ensure consistency.

Evidence attribution
All information in an entry is attributed to its original source e.g. experimental data from a publication.

Gene Ontology (GO) terms
Assigned manually based on experimental data from the literature and are also added based on automatic methods. GO terms are also imported from other GO consortium members.

Publications
Literature sources for the curation of an entry are cited with the type of information that was retrieved.

References

Uniprot is funded by National Institutes of Health, European Molecular Biology Laboratories, Swiss Federal Government, British Heart Foundation, Parkinson’s Disease United Kingdom and National Science Foundation.