

Ontologies for Biology: The Gene Ontology Project

Midori A. Harris, Ph.D.

GO Editor EMBL - EBI
Wellcome Trust Genome Campus Hinxton
Cambridge CB10 1SD UK
Email: midori@ebi.ac.uk
Tel: +44 (0) 1223 494667
Fax: +44 (0) 1223 494468

Introduction

The Gene Ontology (GO) project is a collaborative effort to construct and apply controlled vocabularies, or ontologies, to facilitate the biologically meaningful annotation of genes and their products in a wide variety of databases. Participating groups include the major model organism databases and other database groups such as the UniProt Consortium (Swiss-Prot + TrEMBL + PIR), the Genome Knowledgebase project, The Institute for Genomic Research (TIGR), and others.

The GO project maintains three vocabularies describing different aspects of molecular and cell biology: Molecular function describes activities, such as catalytic or binding activities, at the molecular level. Biological process describes broad objectives, each accomplished by one or more ordered assemblies of molecular functions. Cellular component describes locations where a gene product may act, and includes both subcellular structures and macromolecular complexes.

The GO vocabularies were originally developed for the description of gene products in databases, and many annotation data sets are made available to the public by GO Consortium members. The GO vocabularies and annotations are part of community resource that also includes software tools for working with the ontologies and annotations, project documentation, and links to relevant literature.

The GO project has also provided a model for the development of ontologies for additional aspects of biology. Chief among the more recently developed vocabularies is the Sequence Ontology (SO), which provides a structured controlled vocabulary for sequence annotation, for the exchange of annotation data and for the description of sequence objects in databases. The SO and other emerging shared, structured vocabularies are publicly available from the Open Biology Ontologies web site (<http://obo.sourceforge.net/>). Ontologies must meet five criteria for inclusion in OBO: openness, sharable syntax (such as the GO syntax or OWL), orthogonality to other OBO ontologies, shared ID space, and term definitions.