



Gene Ontology (GO)

Overview



What is an Ontology?

- Formal representation that models a domain of knowledge
- Provide explicit representation of the model
 - Machine processable
 - Standardized
 - Interoperable
 - Reasoning

Ontologies define



Concepts/Classes:

Example:

...

Biological Process

|_ Cellular Process

|_ Biological Regulation

|_ Regulation of Biological Processes

|_ Regulation of Signaling Process

|_ Regulation of Cellular Process

...

...

...

Properties/Relations

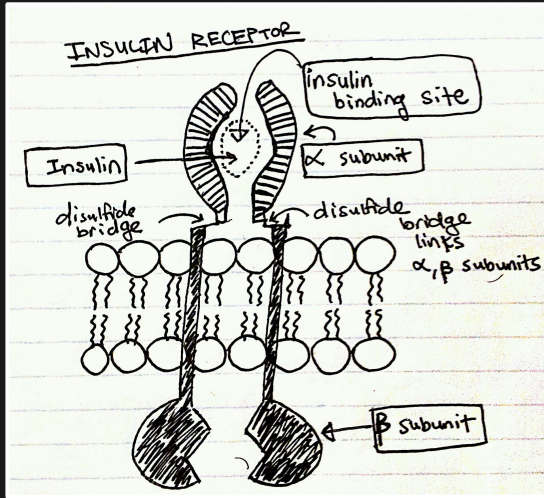
- Object properties: Between 2 objects
- Ex: is part of, regulates, ...

mitochondrial membrane **is part of** mitochondrial envelope
latency-replication decision **regulates** release from viral latency

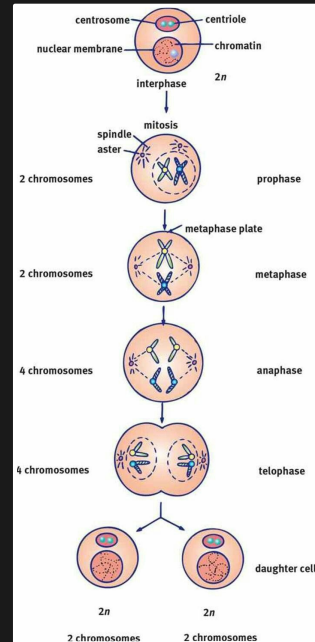
- Data properties: Attributes

Gene Ontology - 3 areas of focus

1) molecular function

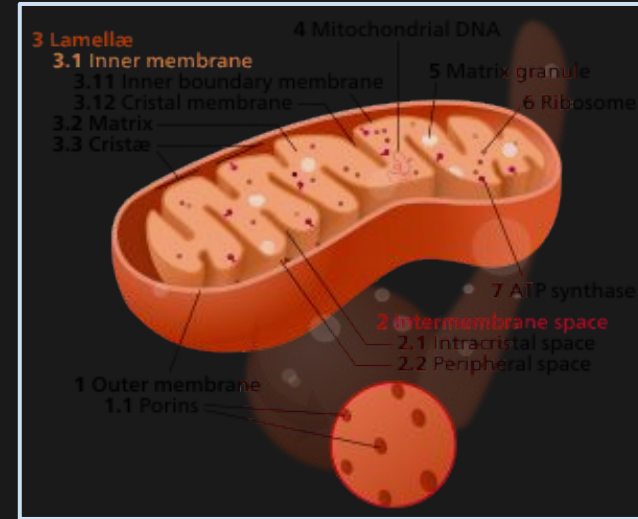


2) biological process



3) cellular component

Location of gene



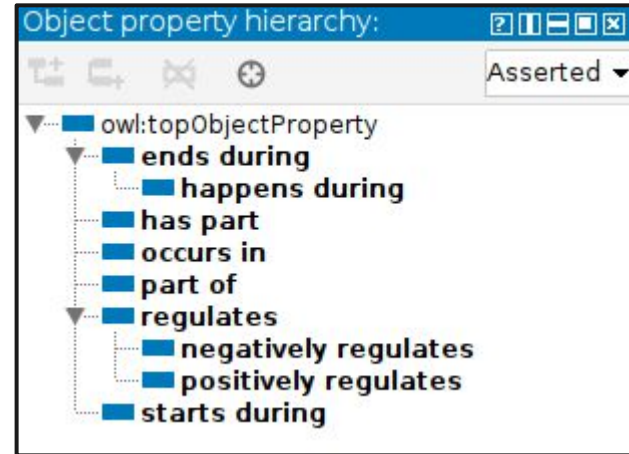
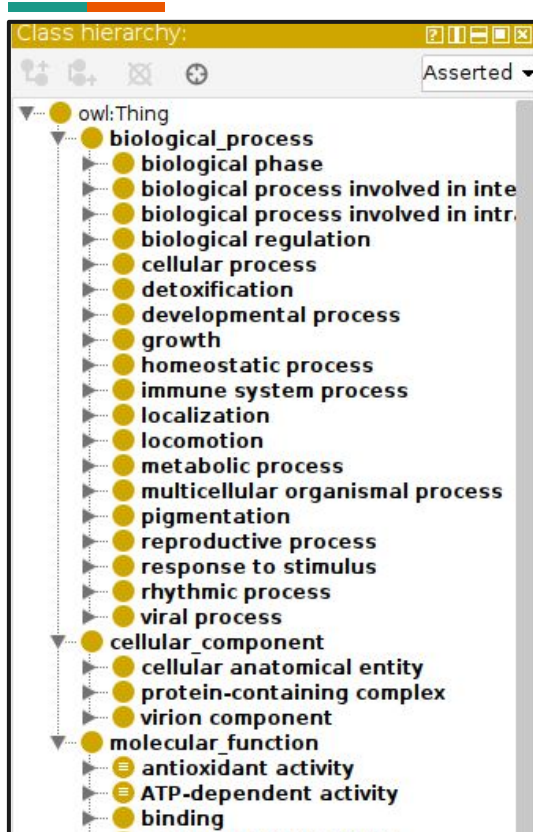
https://en.wikipedia.org/wiki/Mitochondrial_matrix

Core Definitions



- **Molecular function:** “A molecular process that can be carried out by the action of a single macromolecular machine, usually via direct physical interactions with other molecular entities. Function in this sense denotes an action, or activity, that a gene product (or a complex) performs.”
- **Biological Process:** “A biological process is the execution of a genetically-encoded biological module or program. It consists of all the steps required to achieve the specific biological objective of the module. A biological process is accomplished by a particular set of molecular functions carried out by specific gene products (or macromolecular complexes), often in a highly regulated manner and in a particular temporal sequence.”
- **Cellular component:** “A location, relative to cellular compartments and structures, occupied by a macromolecular machine. There are three types of cellular components described in the gene ontology: (1) the cellular anatomical entity where a gene product carries out a molecular function (e.g., plasma membrane, cytoskeleton) or membrane-enclosed compartments (e.g., mitochondrion); (2) virion components, where viral proteins act, and (3) the stable macromolecular complexes of which gene product are parts (e.g., the clathrin complex).”

GO Ontology



No data properties are defined

Ontology Stats



Metrics

Axiom	576,954
Logical axiom count	92,347
Declaration axioms count	51,622
Class count	51,550
Object property count	9
Data property count	0
Individual count	0
Annotation Property count	64

Class axioms

SubClassOf	81,756
EquivalentClasses	10,551
DisjointClasses	30
GCI count	0
Hidden GCI Count	10,551

Object property axioms

SubObjectPropertyOf	3
EquivalentObjectProperties	0
InverseObjectProperties	1
DisjointObjectProperties	0

Inferred Classes

Class restrictions

polyketide synthase activity

- transferase activity
- cytoskeletal motor activity
- electron transfer activity
- fusogenic activity
- general transcription initiation factor activity
- membrane bending activity
- membrane destabilizing activity
- membrane grommet activity
- molecular adaptor activity
- molecular carrier activity
- molecular function regulator activity
- molecular sequestering activity
- molecular tag activity
- molecular template activity
- molecular transducer activity
- nutrient reservoir activity
- none forming activity

PKS activity

term tracker item [type: xsd:anyURI]
<https://github.com/geneontology/go-ontology/issues/7340>

Description: polyketide synthase activity

Equivalent To +
● 'catalytic activity' and ('part of' some 'polyketide biosynthetic process')

SubClass Of +
● 'catalytic activity'
● 'part of' some 'polyketide biosynthetic process'

General class axioms +

Existential Restrictions
(some)

negative regulation of oxidoreductase activity

- negative regulation of oxidoreductase activity
- negative regulation of oxidoreductase activity

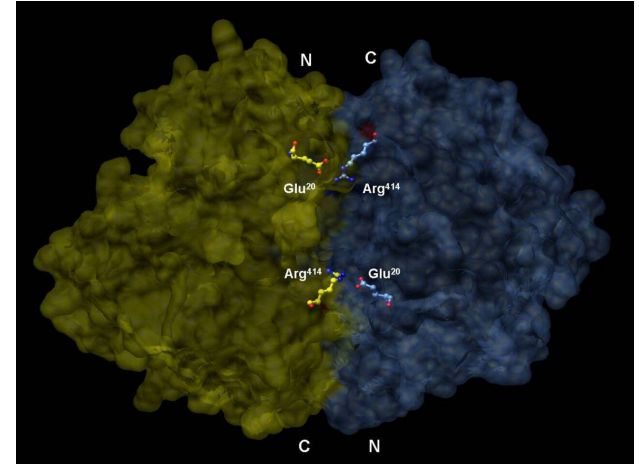
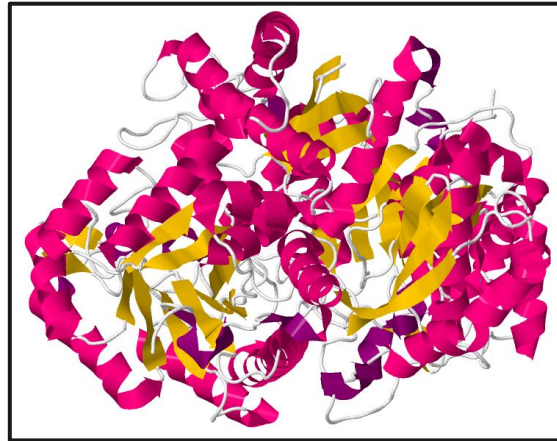
negative regulation of oxidoreductase activity' SubClassOf 'negatively regulates' some 'oxidoreductase activity'

negative regulation of oxidoreductase activity' EquivalentTo 'biological regulation' and ('negatively regulates' some 'oxidoreductase activity')

Let's look at an example

Enolase Annotation (aka **Phosphopyruvate hydratase**)

Click to see annotations



[https://en.wikipedia.org/wiki/Enolase#/media/File:Enolase with differentiated subunits.jpg](https://en.wikipedia.org/wiki/Enolase#/media/File:Enolase_with_differentiated_subunits.jpg)

GO Class - phosphopyruvate hydratase complex

The screenshot displays the GO (Gene Ontology) browser interface. The browser window title is "go (http://purl.obolibrary.org/obo/go/releases/2024-09-08/go.owl) : [/home/uskudarli/research/proteins/ontologies/go.owl]". The address bar shows the URL "http://purl.obolibrary.org/obo/go/releases/2024-09-08/go.owl". The breadcrumb trail is "cellular_component > protein-containing complex > catalytic complex > phosphopyruvate hydratase complex".

The left sidebar shows the "Class hierarchy" for "phosphopyruvate hydratase complex". The hierarchy is as follows:

- peptidase complex
- phenylalanine-tRNA ligase complex
- phosphatase complex
- phosphoenolpyruvate-dependent suc
- phosphoantithenocysteine decar
- phosphopyruvate hydratase complex
- phosphoribosylaminimidazole carbo
- polyketide synthase complex
- proteasome core complex
- proton-transporting ATP synthase coi
- pyrophosphate-dependent phospho
- pyrophosphate-dependent phospho
- pyrophosphate-dependent phospho
- ribonuclease H2 complex
- ribulose biphosphate carboxylase cc
- riboptosome
- serine-tRNA ligase complex
- sulfite reductase complex (NADPH)
- sulfolpyruvate decarboxylase complex
- SUMO activating enzyme complex
- telomerase holoenzyme complex
- thioglucohydrolase complex
- transferase complex
- tricarboxylic acid cycle heteromeric er
- trRNA-specific adenosine-34 deaminas
- trRNA-splicing ligase complex
- ubiquinone biosynthesis complex
- ubiquitin activating enzyme complex
- urease complex
- viral terminase, large subunit
- xylanosome
- Cdc24p-Far1p-Gbetagamma complex
- chloroplast thylakoid membrane protein
- CNTRF-CLCF1 complex
- cohesin complex
- collagen trimer
- complement component C1 complex
- complement component C1q complex
- condensin complex
- counting factor complex
- CRLF-CLCF1 complex

The main panel shows the details for the class "phosphopyruvate hydratase complex". The "Annotations" tab is selected, showing the following annotations:

- rdfs:label** phosphopyruvate hydratase complex
- id** [GO:0000015](#)
- has_obo_namespace** cellular_component
- definition** A multimeric enzyme complex, usually a dimer or an octamer, that catalyzes the conversion of 2-phospho-D-glycerate to phosphoenolpyruvate and water.
- database_cross_reference** [GOC:jl](#)
- database_cross_reference** [ISBN:0198506732](#)
- has_exact_synonym** enolase complex
- in_subset** [goslim_metagenomics](#)

The "Description" tab is also visible, showing the description: "phosphopyruvate hydratase complex".

The "SubClass Of" section shows the following subclasses:

- 'catalytic complex'
- 'part of some cytosol'

The "General class axioms" section is empty.

The "SubClass Of (Anonymous Ancestor)" section is empty.

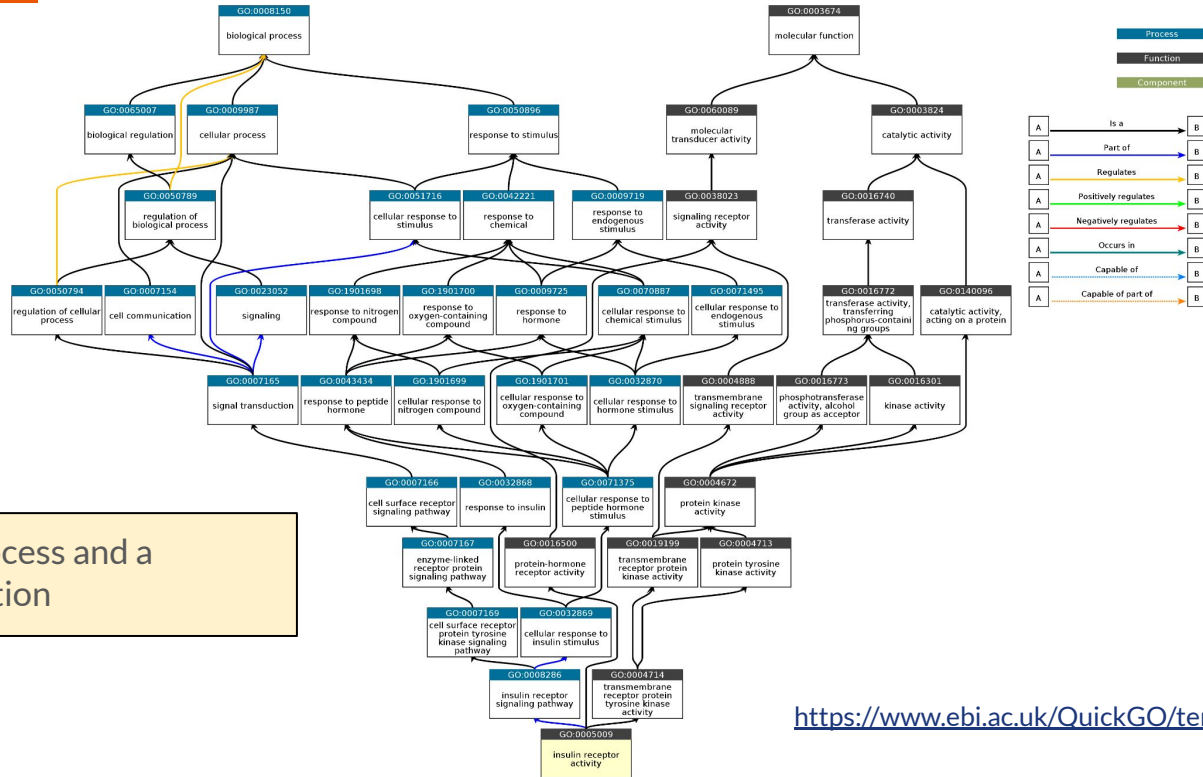
The "Instances" section is empty.

The "Target for Key" section is empty.

The "Disjoint With" section is empty.

The bottom status bar indicates "To use the reasoner click Reasoner > Start reasoner" and "Show Inferences".

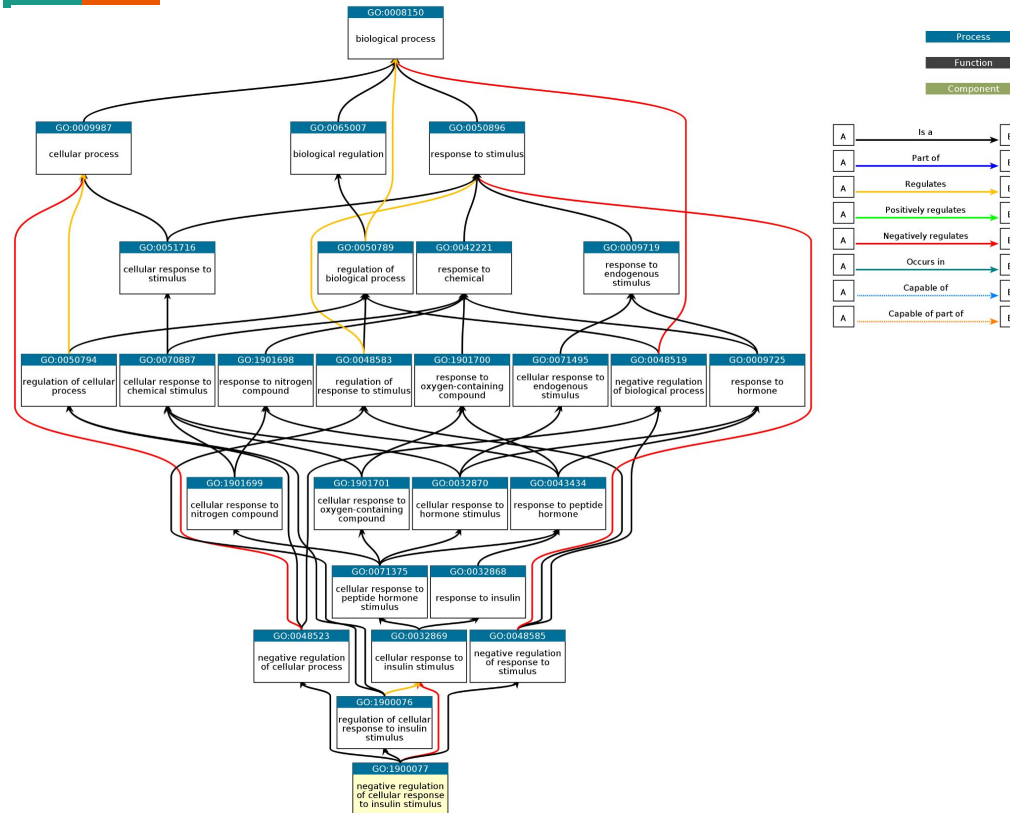
Ancestor Chart of GO:0005009 (Insulin Receptor Activity)



A biological process and a molecular function

<https://www.ebi.ac.uk/QuickGO/term/GO:0005009>

More complex regulation functions – GO:1900077 (negative regulation of cellular response to insulin stimulus)



<https://www.ebi.ac.uk/QuickGO/term/GO:1900077>

GO Annotations



4 pieces of required information to uniquely identify a **GO annotation**.

- Gene product (may be a protein, RNA, etc.)
- GO term
- Reference
- [Evidence](#)

Always gives evidence related with assertions.

[Also](#), additional components information/[relations](#) may be defined, the above are essential.

Semantics of a GO annotations



GO terms are about gene products

The statements that they describe:

- Molecular Function: the molecular activities of individual gene products
- Cellular Component: where the gene products are active
- Biological Process: the pathways and larger processes to which that gene product's activity contributes

GO Annotations Principles



- represent the normal functions of gene products.
- A gene product can be annotated to zero or more terms from each ontology.
- Supported by [GO Evidence Codes](#) from the [Evidence and Conclusions Ontology](#) and a reference.
- Gene products are annotated to the **most granular** term in the ontology that is supported by the available **evidence**.
- By the transitivity principle, an annotation to a GO term is true for all its parents (except for *NOT* annotations).
- GO annotations are meant to reflect the most up-to-date view of a gene product's role in biology.
- Annotations for a given gene product may change to reflect changes in knowledge and/or the ontology.
- There is an **open-world** assumption, that is, if a gene product is unannotated then its role is still unknown.

NOT annotations



NOT is used when a GO term is **expected to apply** to a gene product, but an experiment, sequence analysis, etc. **proves otherwise**.

NOT makes an **explicit statement** that a gene product has been experimentally demonstrated **not to be able to carry out a particular activity or it has been shown to have lost that function** (e.g. sequence analysis showing a loss of an active site or rapid divergence after a duplication event) over the course of evolution.

Uses of GO Terms



- [UniProt-Gene Ontology Annotation \(UniProt-GOA\)](#) - Manual and electronic annotation of proteins curated by European Bioinformatics Institute
- [RHEA](#) - expert-curated knowledge of chemical and transport reactions of biological interest - and the standard for enzyme and transporter annotation in UniProtKB
- [University College London Functional Gene Annotation](#) - Manual annotation of human proteins and microRNAs involved in cardiovascular and dementia-relevant processes
- And many many more



References

1. Gene Ontology docs @ <https://geneontology.org/docs/>
2. The related ontologies @ <https://geneontology.org/docs/download-ontology/>
3. [UniProtKB](#)
- 4.