

# The GOA Database

André Schuiteman  
LIACS  
28 October 2004



## GOA = Gene Ontology Annotation

<http://www.ebi.ac.uk/GOA/>

Aims of this database:

- To provide annotations to the UniProt Knowledgebase
- Using the vocabulary of the Gene Ontology (GO) Consortium

## Gene Ontology resource

<http://www.geneontology.org/>



A dynamic controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing.

Three separate ontologies: molecular function, biological process and cellular component, to describe gene products.

Each vocabulary is structured as directed acyclic graphs (DAGs).

Currently more than 17,000 terms included.

These terms are to be used as attributes of gene products by collaborating databases, facilitating uniform queries across them.

## AmiGO

### RNA interference, siRNA loading onto RISC

Accession: GO:0035087

Aspect: process

Synonyms: None

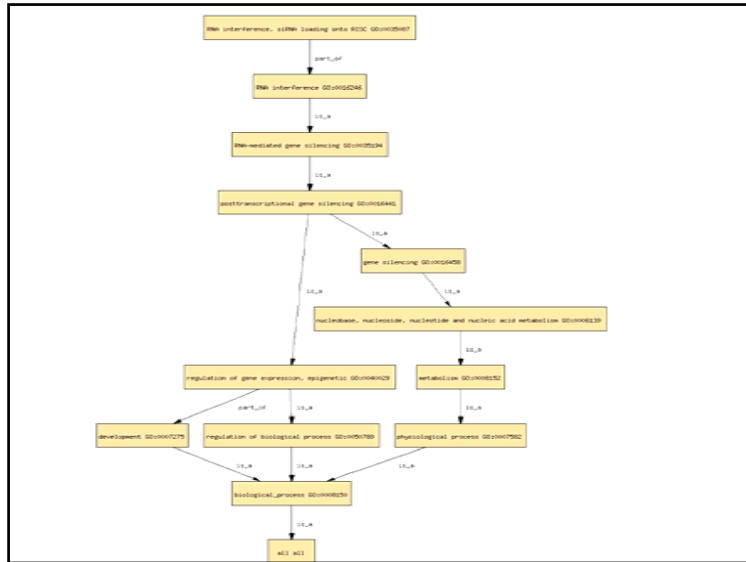
Definition:


The transfer of small interfering RNA molecules (siRNAs) from the Dicer family of enzymes that cleave the double-stranded RNA, onto the nuclease-containing

#### Term Lineage

all (153306)  
 GO:000150 : biological process (103163)  
 GO:0007275 : development (15301)  
 GO:0040029 : regulation of gene expression, epigenetic (600)  
 GO:0016441 : posttranscriptional gene silencing (79)  
 GO:0035194 : RNA-mediated gene silencing (65)  
 GO:0016246 : RNA interference (65)  
 GO:0035087 : RNA interference, siRNA loading onto RISC (2)  
 GO:0007682 : physiological process (65343)  
 GO:0000152 : metabolism (4000)  
 GO:0006139 : nucleobase, nucleoside, nucleotide and nucleic acid metabolism (15839)  
 GO:0016488 : gene silencing (313)  
 GO:0016441 : posttranscriptional gene silencing (79)  
 GO:0035194 : RNA-mediated gene silencing (65)  
 GO:0016246 : RNA interference (65)  
 GO:0035087 : RNA interference, siRNA loading onto RISC (2)  
 GO:0050789 : regulation of biological process (13216)  
 GO:0040029 : regulation of gene expression, epigenetic (600)  
 GO:0016441 : posttranscriptional gene silencing (79)  
 GO:0035194 : RNA-mediated gene silencing (65)  
 GO:0016246 : RNA interference (65)  
 GO:0035087 : RNA interference, siRNA loading onto RISC (2)

Graphical View



  
 the universal protein resource

**UniProt**  
<http://www.ebi.uniprot.org/index.shtml>

The UniProt Knowledgebase (UniProt) is the central access point for curated protein information, including function, classification, and cross-reference. Accessible via text search, BLAST similarity search, and FTP.


UniProt Release 3.0 consists of 1,612,609 entries (UniProt/Swiss-Prot: 163,235 entries and UniProt/TrEMBL: 1,449,374 entries).

Warehouse  
 Prediction Search  
 InterPro Search  
 CluST Search  
 Entry List Search  
 Data Set Manager  
 BLAST  
 FAQ  
 Help Desk  
 Download

Your Query Result Sets (Page - 1) [Data Set Manager]

Query: FullText: "cytochrome" [Swiss-Prot & TrEMBL]  
 has 69741 entries

ID/Accession	Protein Name	Length	Organism Name	Taxon Group
SCE2_YEAST / P03078	Intron-encoded DNA endonuclease a14 precursor (DNA endonuclease 1-ScEl1) [Contains: Truncated, nonfunctional cytochrome oxidase 1; DNA endonuclease a14 (EC 3.1.-.-) (Intron-encoded endonuclease 1-ScEl1)].	556	Saccharomyces cerevisiae (Baker's yeast)	Eukaryota Fungi
OX11_SCHPO / O14300	Inner membrane protein oxa1-1, mitochondrial precursor (Cytochrome oxidase biogenesis protein 1-1) (Sp1).	374	Schizosaccharomyces pombe (Fission yeast)	Eukaryota Fungi
OX12_SCHPO / O43092	Inner membrane protein oxa1-2, mitochondrial precursor (Cytochrome oxidase biogenesis protein 1-2) (Sp2).	409	Schizosaccharomyces pombe (Fission yeast)	Eukaryota Fungi
OXA1_ARATH / Q42191	Inner membrane protein OXA1, mitochondrial precursor (Oxidase assembly 1 protein) (AOX1).	429	Arabidopsis thaliana (Mouse-ear cress)	Eukaryota Vridiplantae
OXA1_HUMAN / Q15070	Inner membrane protein OXA1, mitochondrial precursor (Oxidase assembly 1-like protein) (OXA1-like protein) (OXA1H1) (H9a).	435	Homo sapiens (Human)	Eukaryota Metazoa
OXAL_MOUSE / Q8BGA9	Inner membrane protein OXA1L, mitochondrial precursor (Oxidase assembly	433	Mus musculus (Mouse)	Eukaryota Metazoa

  
 the universal protein knowledgebase

Home > Database > UniProt Protein Viewer  
 Searched by:  Text Search UniProt Knowledgebase

**Basic UniProt Protein Viewer**

Your Query Result Sets (Page - 1) [Data Set Manager]

Protein SCE2\_YEAST  
 P03078 | O14300 | O43092 | Q42191 | Q15070 | Q8BGA9 | Q02339 | Q35798 | Q927X2

Basic | Extended  
 Viewers: Facts | Flat File | XML | ExPASy | SRS | PIR

**General information about the UniProt/Swiss-Prot entry**

Entry name	SCE2_YEAST
Primary accession number	P03078
Secondary accession numbers	Q02339 Q35798 Q927X2
Entered in Swiss-Prot	Release 01, 21-JUL-1996
Sequence was last modified	Release 45, 25-OCT-2004
Annotations were last modified	Release 45, 25-OCT-2004

**Protein description**

Protein name	Intron-encoded DNA endonuclease a14 precursor
--------------	---

Protein description		
Protein name	Intron-encoded DNA endonuclease a14 precursor	
Synonyms	DNA endonuclease I-SceII	
Contains	Protein name	Truncated, nonfunctional cytochrome oxidase 1
	Protein name	DNA endonuclease a14
	Enzyme classification	EC 3.1.-.-
	Synonyms	Intron-encoded endonuclease I-SceII
Origin of the protein		
Gene	Gene name  a14 Synonyms  ENS2, I-SCEII Locus names  Q0065	
From	Saccharomyces cerevisiae (Baker's yeast)[TaxID:4932]	
Taxonomy	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces	
Encoded on	Mitochondrion	
References		
[1]	SEQUENCE FROM N.A., AND VARIANT 457-LYS-THR-458. STRAIN=O273-108. MEDLINE=6106985; PubMed=6254986; [NCBI, ExPASy, EBI, Israel, Japan] Bonitz S.G., Coruzzo G., Thalmeier B.E., Teagloff A., Macino G.: "Assembly of the mitochondrial membrane system. Structure and nucleotide sequence of the gene coding for subunit 1 of yeast cytochrome oxidase.", J. Biol. Chem. 255:11927-11941(1980).	
[2]	SEQUENCE OF 245-531 FROM NUCLEIC ACID, MUTAGENESIS OF GLU-362, AND MRNA MATURASE ACTIVITY. MEDLINE=82245571; PubMed=6285204; [NCBI, ExPASy, EBI, Israel, Japan] Dourdin G., Jazic G., Stommes R.P.: "Single base substitution in an intron of oxidase gene compensates splicing defects of the cytochrome b gene.",	

Comments	
FUNCTION	Mitochondrial DNA endonuclease involved in intron homing. It introduces a specific double-strand break at the junction of the two exons a1-35 of the COX1 gene and thus mediates the insertion of an intron, containing its own coding sequence (group 1 intron), into an intronless gene. Recognizes with limited specificity and cleaves the sequence 5'-TTTGTACACCTGAAGTA-3'. The protein may acquire mRNA maturase activity, like the closely related b14, through a single amino acid substitution Glu-362 to Lys or when present together with a mutant form of the imported mitochondrial leucyl-tRNA synthetase NAM2.
COFACTOR	Magnesium ion or Manganese(II) ion.
SUBUNIT	Homodimer (Probable).
SUBCELLULAR LOCATION	Mitochondrial.
PTM	The mature protein may arise from proteolytic cleavage of an in-frame translation of COX1 exons 1 to 4 plus intron 4, containing the a14 open reading frame. Cleavage would take place close to the Met-299 resulting in an active endonuclease of about 30 kDa.
MISCELLANEOUS	Residues 299 to 556 are sufficient for endonuclease and intron homing activity.
SIMILARITY	In the C-terminal section; belongs to the LAGLIDADG endonuclease family.
CAUTION	Ref.1 sequence differs from that shown because COX1 was not predicted to be expressed alternatively as a fusion with intron 4.
Copyright	
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> )	
Cross-references	
EMBL	V00694; CA24070.1; ALT_SEQ [EMBL/GenBank/DBJ] [CoDingSequence] V00694; CA24064.1; ALT_SEQ [EMBL/GenBank/DBJ] [CoDingSequence] V00703; CA24076.1; ~ [EMBL/GenBank/DBJ] [CoDingSequence] S76641; A082126.1; ~ [EMBL/GenBank/DBJ] [CoDingSequence]

PIK	SwissProt; SwissProt; 578649; QXBY34.				
GermOnline	144523; ~.				
REBASE	2616; I-SceII.				
SGD	S000007264; A14.				
InterPro	IPR001982; Endonuc_LAG/HNH; Graphical view of the domain structure				
Pfam	PF00991; LAGLIDADG_1; 2. Pfam graphical view of domain structure				
PROSITE	PS50855; COX1; 1. PROSITE graphical view of domain structure				
Keywords					
Endonuclease					
Hydrolase					
Intron homing					
Mitochondrion					
mRNA processing					
mRNA splicing					
Nuclease					
Features					
Type	From	To	Length	Description	Feature ID
CHAIN	?	556		DNA endonuclease a14.	
CHAIN	1	?		Truncated, nonfunctional cytochrome oxidase 1.	
DOMAIN	1	240	240	COX1 exon 1 to 4 encoded.	
DOMAIN	241	556	316	COX1 intron 4 encoded.	

CONFLICT	315	315		M -> I (in Ref. 3).
CONFLICT	330	330		M -> I (in Ref. 3).
<a href="#">FeatureTableViewer</a> <a href="#">FeatureAligner</a>				
Sequence information				
Length	556 AA			
Molecular weight	63311 Da			
CRC64	4FD968F23DABA7A [This is a checksum on the sequence]			
<pre> MYORVLYSTN AKDIAYLYFH LAIFSGMAGT ANGLIIRLEL AAPGGQYING 50 NSQLFNVLVY GHAVLMIFFL VMPALLOGFC NYLLPLMIGA TDTAFPRINN 100 IAFFWFLMGL WCLVTSTLVE SGAGTGVTVY PPLSSIOAGS GPSVSLAIFA 150 LHLTSGISLL GAINFIVTL NWRWGWTHR KPLFLFVWGF ITAFLLLSGL 200 PVLSAGITML LIDRNFTYF FEVSGGGDFI IYERHLFVFG QTVATTIMLM 250 MYNDHFSKC WLLKRWITN IMSTLFLKLF VMKPKSYNNO QDQNNNTML 300 KKDINKRSE TTEKMLNNSM NKKFNQWLAG LIDGQYVGI VSKKYVSLRI 350 TVALEDEAL KEIQNKFGS IKLRSGYKAI RYRLTKWTKM IKLINAWKN 400 IRNTKRLVDF NKVCILLID FIYPIKLTQD NSWFVGFDDA DGTINYSFEN 450 NRPQLTISVT NKYLQWQYF KNILGNIYF DKSQNGYVW SIOSKDMVLR 500 FINDYIKNPF SKYTKMKNLY LSKFEPNLE LKAYNKSSDQ NQYKAVLNFE 550 NKVKNK </pre>				
Pop-Up FASTA View <span style="float: right;">UniParc   UniRef100   UniRef90   UniRef50</span>				

EMBL-EBI European Bioinformatics Institute

GOA What can I do with GOA?

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

- access functional information for the human proteome (GOA Human) or for any protein in EBI's protein databases (GOA-UniProt)
- ask questions such as "find me all proteins involved in apoptosis(GO:0006915) but not involved via death domain receptors (GO:0008625), and then find their coding sequences".
- use [GO-Slim](#) to summarize the biological attributes of a proteome, compare proteomes, or find out what proportion of a proteome is involved in eg. "transport".
- incorporate our manual annotation into your own databases to enhance your dataset, or use it to validate your automated way of deriving information about gene function.
- map GO terms to your own datasets; for example, our GO mapping to InterPro entries ([InterPro2GO](#)) can be used to annotate mass spectrometry or microarray data.
- find the location of human genes mapped to a particular GO term using [ensembl GO-View](#)

For information, comments and/or suggestions, please email us at [goa@ebi.ac.uk](mailto:goa@ebi.ac.uk)

Note: The current release has 5162144 entries.

EMBL-EBI European Bioinformatics Institute

Quick Search Library Page Query Form Tools Results Projects Views Databases HELP

Reset search GOA

**Search Options**

Combine search terms with: & (AND)

Use wildcards:

Get results of type: Entry

**Fields you can search**

In a single field, you can separate multiple values by &, |, !

Your search terms: cytochrome

**Result Display Options**

View results using: GoSimple

or

Create a view

Show 30 results per page

**Create a view**

Select the fields you want displayed in your view and choose the format

Choose 1 or more fields:

Pubmed ID  
GO Evidence Code  
Associated with Accession Number  
Function/Process/Component  
Description  
IP1 ID  
NCBI TaxID

Display As: Table List

Tips Search

EMBL-EBI European Bioinformatics Institute

Quick Search Library Page Query Form Tools Results Projects Views Databases HELP

Reset Query "goa:all[with:cytochrome]" found 530340 entries

**Apply Options for:**

selected results only  
unselected results only

**Result Options**

Show tools relevant to these results: [Tools](#)

Link to related information: [Link](#)

Save results: [Save](#)

**Display Options**

View results using: GoSimple

Show 10 results

GOA	WB	Accession Number	Entry Name	GO ID	Pubmed ID	GO Evidence Code	Associated with accession number	Function/Process/Component	Description	IP1 ID	GB/YFR	NCBI TaxID	Date	Assigned by
GOA:84021	RefSeq	NP_010204541	NP_010204541	GO:0021627	GOA:Interpro	EA			cytochrome P450 family 2, subfamily 11	Q55Q43092	protein	10292	20041007	UniProt
GOA:84022	RefSeq	NP_010204541	NP_010204541	GO:0001118	GOA:Interpro	EA			cytochrome P450 family 2, subfamily 11	Q55Q43093	protein	10292	20041007	UniProt
GOA:45023	RefSeq	NP_010204541	NP_010204541	GO:0018732	GOA:Interpro	EA			cytochrome P450 family 2, subfamily 11	Q55Q43093	protein	10292	20041007	UniProt
GOA:84012	RefSeq	NP_010204541	NP_010204541	GO:0001118	GOA:Interpro	EA			cytochrome P450 inducible isoform	Q55Q43093	protein	3605	20041007	UniProt

The result.