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/

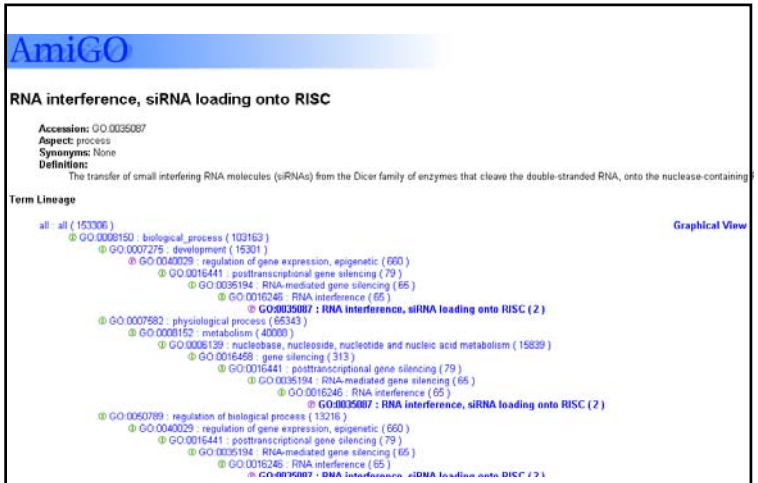
A dynamic controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing.

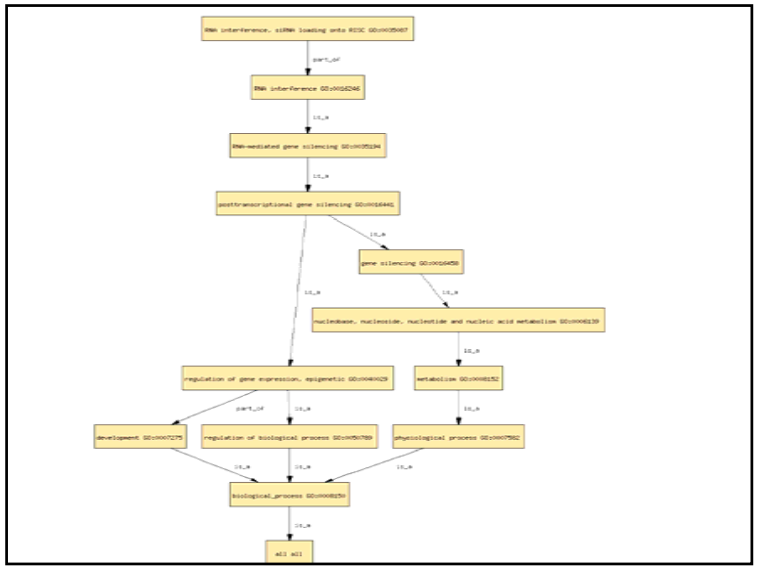
Three separate ontologies: molecular function, biological process and cellular component, to describe gene products.

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.





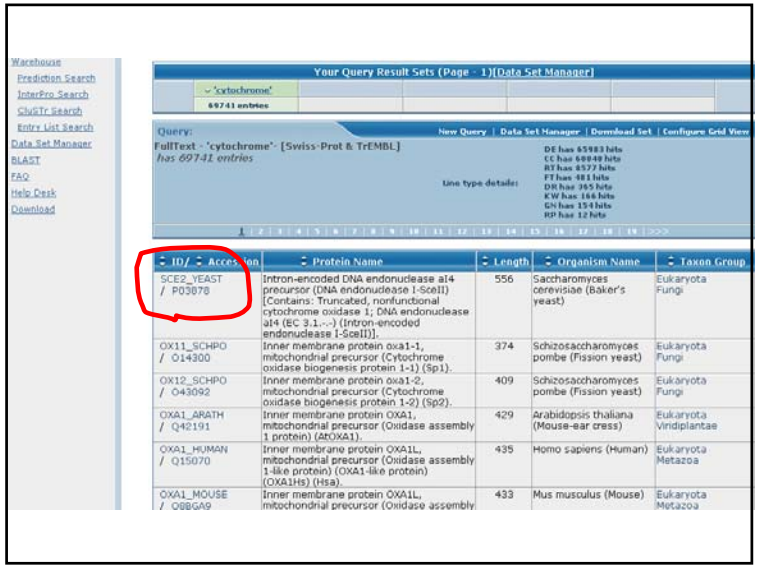
UniProt

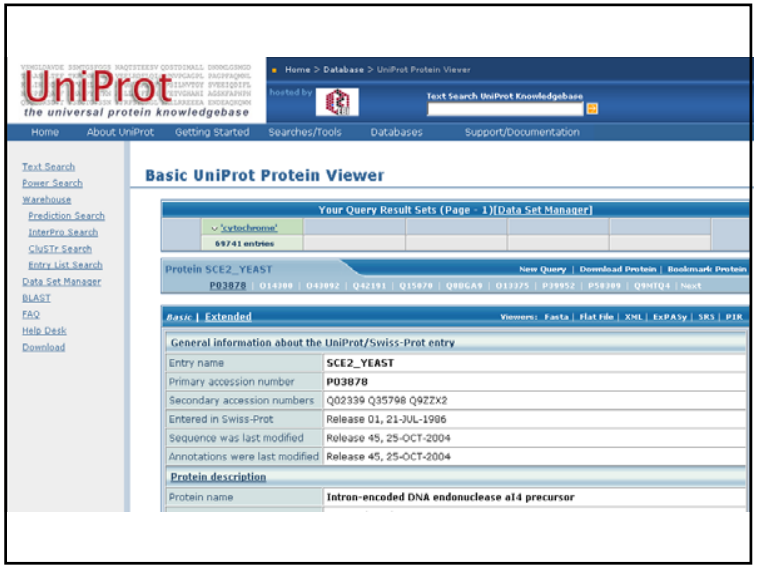
http://www.ebi.uniprot.org/index.shtml

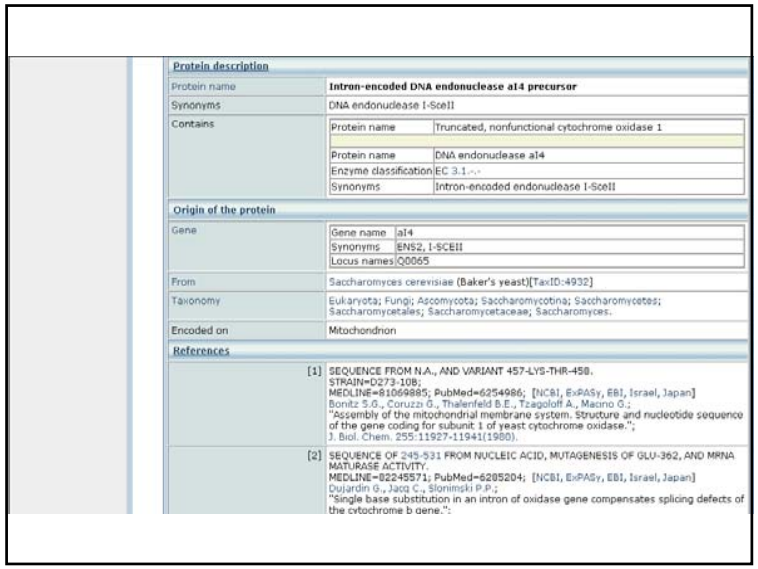
The UniProt Knowledgebase (UniProt) is the central access point for curated protein information, including function, classification, and cross-reference.

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) .







Protein description

Protein name	**Intron-encoded DNA endonuclease aI4 precursor**
Synonyms	DNA endonuclease I-SceII
Contains	Protein name: Truncated, nonfunctional cytochrome oxidase 1 Protein name: DNA endonuclease aI4 Enzyme classification: EC 3.1.-.- Synonyms: Intron-encoded endonuclease I-SceII

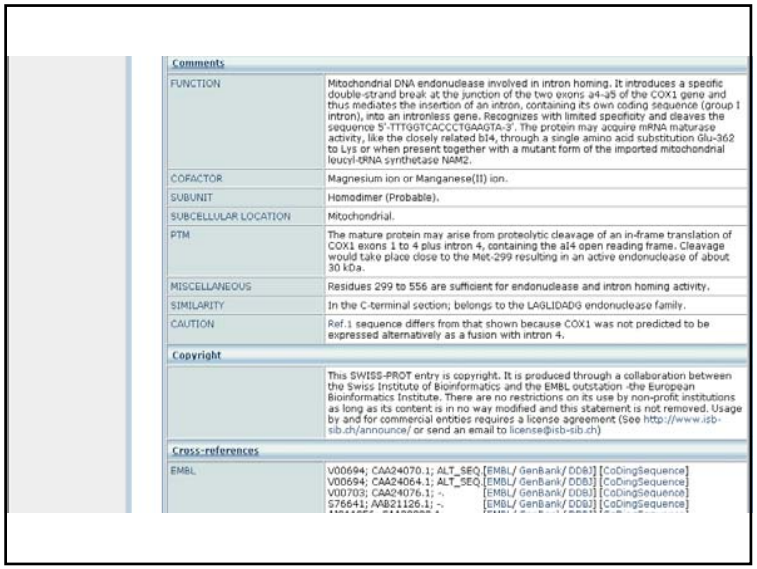
Origin of the protein

Gene	Gene name: aI4 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=D273-10B;
MEDLINE=81069885; PubMed=6254986; [NCBI, ExPASy, EBI, Israel, Japan]
Bonitz S.G., Coruzzi G., Thalenfeld B.E., Tzagoloff 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]
Dujardin G., Jacq C., Slonimski P.P.;
"Single base substitution in an intron of oxidase gene compensates splicing defects of the cytochrome b gene.":



Comments

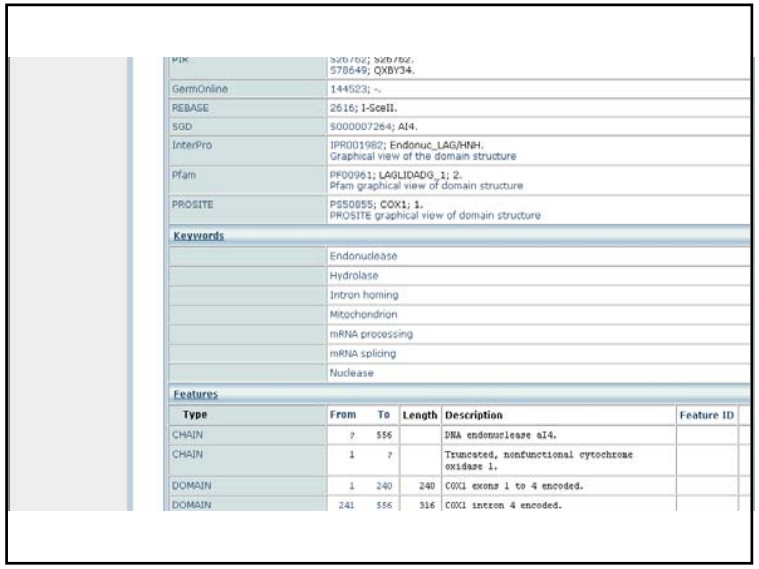
FUNCTION	Mitochondrial DNA endonuclease involved in intron homing. It introduces a specific double-strand break at the junction of the two exons a4-a5 of the COX1 gene and thus mediates the insertion of an intron, containing its own coding sequence (group I intron), into an intronless gene. Recognizes with limited specificity and cleaves the sequence 5'-TTTGGTCACCCTGAAGTA-3'. The protein may acquire mRNA maturase activity, like the closely related bI4, 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 aI4 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 http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch)

Cross-references

EMBL	V00694; CAA24070.1; ALT_SEQ. [EMBL/ GenBank/ DDBJ] [CoDingSequence] V00694; CAA24064.1; ALT_SEQ. [EMBL/ GenBank/ DDBJ] [CoDingSequence] V00703; CAA24076.1; -. [EMBL/ GenBank/ DDBJ] [CoDingSequence] S76641; AAB21126.1; -. [EMBL/ GenBank/ DDBJ] [CoDingSequence]



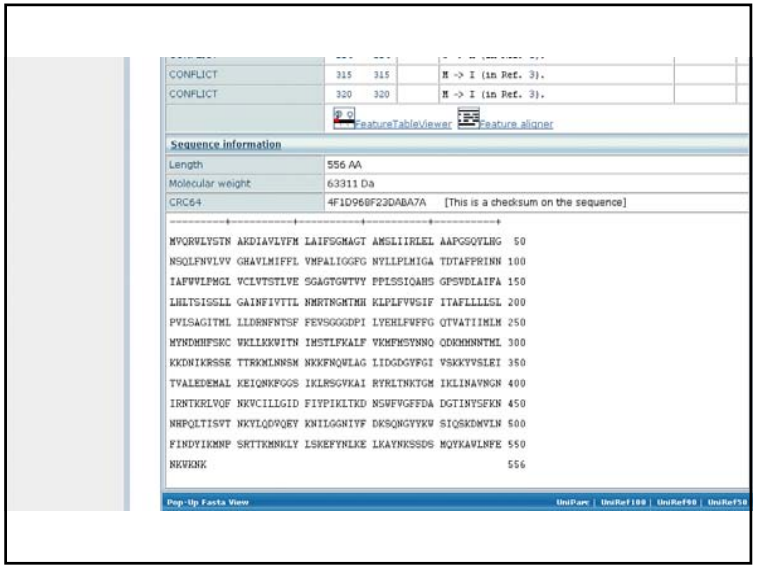
PIR	S26782; S26782.
	S78649; QXBY34.
GermOnline	144523; -.
REBASE	2616; I-SceII.
SGD	S000007264; aI4.
InterPro	IPR001982; Endonuc_LAG/HNH. Graphical view of the domain structure
Pfam	PF00961; LAGLIDADG_1; 2. Pfam graphical view of domain structure
PROSITE	PS50855; COX1; 1. PROSITE graphical view of domain structure

Keywords

Endonuclease
Hydrolase
Intron homing
Mitochondrion
mRNA processing
mRNA splicing
Nuclease

Features

Type	From	To	Length	Description	Feature ID
CHAIN	?	556		DNA endonuclease aI4.	
CHAIN	1	?		Truncated, nonfunctional cytochrome oxidase 1.	
DOMAIN	1	240	240	COX1 exons 1 to 4 encoded.	
DOMAIN	241	556	316	COX1 intron 4 encoded.	



CONFLICT	315	315	M -> I (in Ref. 3).
CONFLICT	320	320	M -> I (in Ref. 3).

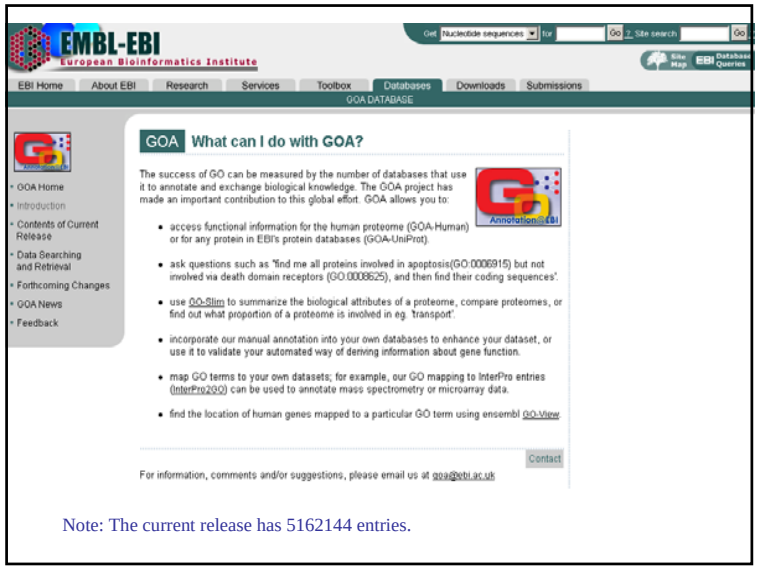
FeatureTableViewer Feature aligner

Sequence information

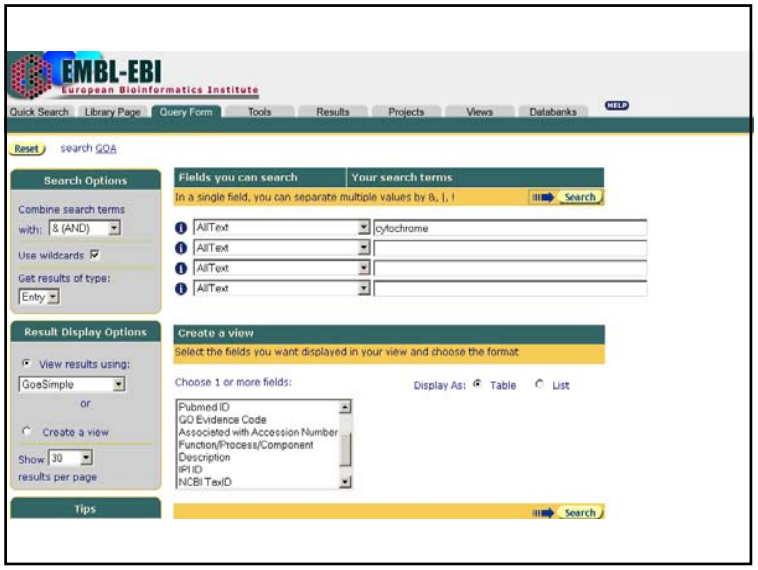
Length	556 AA
Molecular weight	63311 Da
CRC64	4F1D968F23DABA7A [This is a checksum on the sequence]

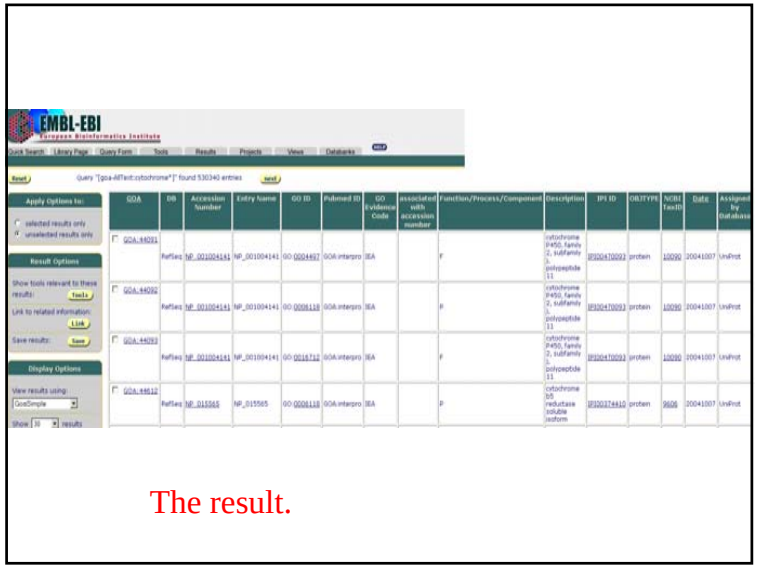
```
MVQRWLYSTN AKDIAVLYFM LAIFSGMAGT AMSLIIRLEL AAPGSQYLHG  50
NSQLFNVLVV GHAVLMIFFL VMPALIGGFG NYLLPLMIGA TDTAFPRINN 100
IAFWVLPMGL VCLVTSTLVE SGAGTGWTVY PPLSSIQAHS GPSVDLAIFA 150
LHLTSIGSLL GAINFIVTTL NMRTNGMTMH KLPLFVWSIF ITAFLLLLSL 200
PVLSAGITML LLDRNFNTSF FEVSGGGDPI LYEHLFWFFG QTVATIIMLM 250
MYNDMHFSKC WKLLKKVITN IMSTLFKALF VKMFMSYNNQ QDKMNNNTML 300
KKDNIKRSSE TTRKMLNNSM NKKFNQWLAG LIDGDGYFGI VSKKYVSLRI 350
TVALEDEMAL KEIQNKFGGS IKLRSGVKAI RYRLTNKTGM IKLINAVNGN 400
IRNTKRLVQF NKVCILLGID FIYPIKLTKD NSWFVGFFDA DGTINYSFKN 450
NHPQLTISVT NKYLQDVQEY KNILGGNIYF DKSQNGYYKV SIQSKDMVLN 500
FINDYIKMNP SRTTKMNKLY LSKEFYNLKE LKAYNKSSDS MQYKAWLNFE 550
NKVKNK                                                  556
```

Pop-Up Fasta View | UniParc | UniRef100 | UniRef90 | UniRef50



Note: The current release has 5162144 entries.





The result.