A Tutorial on Protein Ontology Resources

April, 2010

Cecilia Arighi, Ph.D
arighi@dbi.udel.edu
Research Assistant Professor
Protein Information Resource
Protein Ontology Consortium

Exploring PRO Resources

The Protein Ontology is a controlled vocabulary of terms to describe protein classes and their relationships. PRO provides a means to refer to a specific protein object. Each protein class has a distinct PRO ID, therefore a modified and unmodified form of given protein are two objects in the ontology. In this tutorial, you will learn how to navigate the PRO website, download the ontology and annotations (PAF).

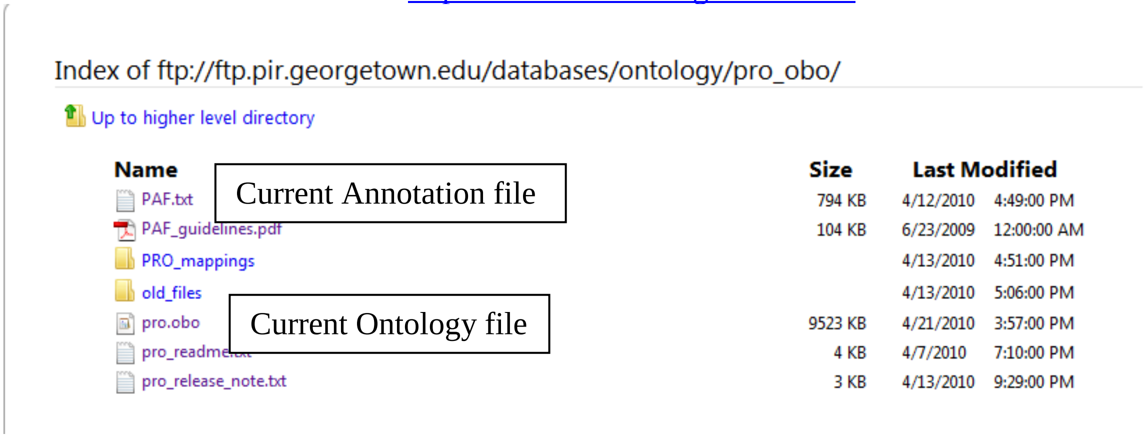
Downloading the ontology and annotations

If you want to incorporate PRO ontology or annotations in your computation work, you may need to download them.

Go to the PRO site: pir.georgetown.edu/pro/
In the left menu, select Downloads to be directed to the ftp site.

The Current Ontology is available at:
ftp://ftp.pir.georgetown.edu/databases/ontology/pro_obo/

PRO file is in OBO 1.2 format and should be opened with OBO Edit 1.1 or higher. This editor can be downloaded from http://www.oboedit.org/index.html.



Index of ftp://ftp.pir.georgetown.edu/databases/ontology/pro_obo/

Up to higher level directory

Name	Size	Last Modified
PAF.txt	794 KB	4/12/2010 4:49:00 PM
PAF_guidelines.pdf	104 KB	6/23/2009 12:00:00 AM
PRO_mappings		4/13/2010 4:51:00 PM
old_files		4/13/2010 5:06:00 PM
pro.obo	9523 KB	4/21/2010 3:57:00 PM
pro_readme	4 KB	4/7/2010 7:10:00 PM
pro_release_note.txt	3 KB	4/13/2010 9:29:00 PM

The ontology is also available through the:
<u>OBO Foundry</u>: http://www.obofoundry.org/ in OBO and OWL formats.
<u>Bioportal</u>: http://bioportal.bioontology.org/

Note the flat-file format:

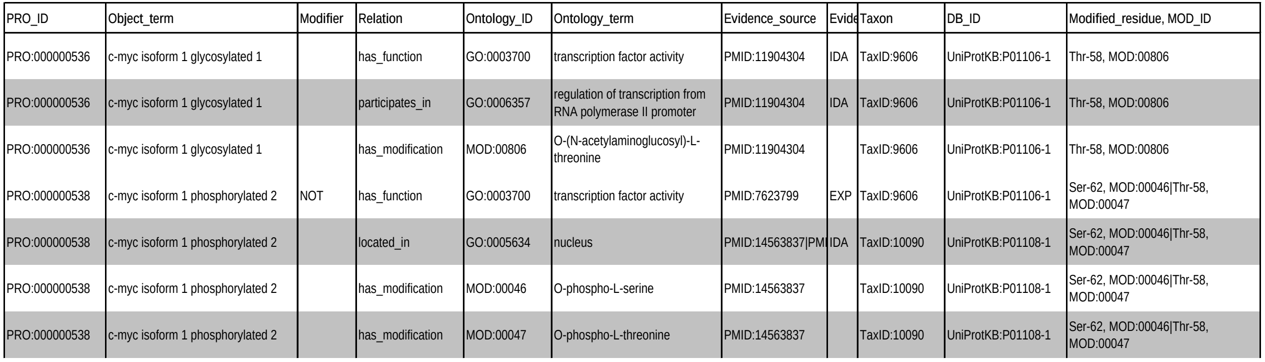
```
format-version: 1.2
date: 02:04:2009 14:51
saved-by: arighic
auto-generated-by: OBO-Edit 1.101
default-namespace: pro
remark: release: 5.0, version 1
```

```
[Term]
id: PRO:000000001
name: protein
def: "A biological macromolecule that is composed of amino
acids linked in a linear sequence (a polypeptide chain) and
is genetically encoded. Proteins descended from a common
ancestor can be classified into families and superfamilies
composed of products of evolutionarily-related genes. The
domain architecture of a protein is described by the order of
its constituent domains. Proteins with the same domains in
the same order are defined as homeomorphic." [PRO:WCB]

[Term]
id: PRO:000000002
name: E3 ubiquitin ligase SFC complex, Skp1 subunit
def: "A protein with a core domain composition consisting of
an N-terminal Skp1 family, tetramerisation domain (PF03931)
followed by a Skp1 family, dimerization domain (PF01466).
Skp1 proteins bind several F-box-containing proteins, and are
involved in the ubiquitin protein degradation pathway."
[PRO:CNA]
comment: Category=family.
xref: PIRSF:PIRSF028729
is_a: PRO:000000001 ! protein

[Term]
id: PRO:000000003
name: HLH DNA-binding protein inhibitor
def: "A protein with a core domain composition consisting of
an Helix-loop-helix DNA-binding domain (PF00010) (HLH),
common to the basic HLH family of transcription factors, but
lacking the DNA binding domain to the consensus E box
response element (CANNTG). By binding to basic HLH
transcription factors, Id proteins regulate gene expression."
[PRO:CNA]
comment: Category=family.
synonym: "DNA-binding protein inhibitor ID" EXACT []
synonym: "ID protein" RELATED []
xref: PIRSF:PIRSF005808
is_a: PRO:000000001 ! protein
```

After some version information, there is a stanza of information about each term. The definition is tied to the id. Notice that each definition is followed by [ ] in which the source of the definition is provided. The source can be a database identifier, or a curator in which case we provide his initials, e.g. [PRO:CNA]

Annotations: The annotations to PRO are in the PAF.txt distribution file as indicated above. Open the README file and the PAF guidelines.pdf to learn about the structure of this file. PAF is a tab delimited file that tries to follows gene ontology association (GAF) file as much as possible to facilitate interoperability. The object of annotation is a PRO term. However, PRO annotates the PRO term using others ontologies on top of GO. Then the PAF file has to accommodate for this requirement.

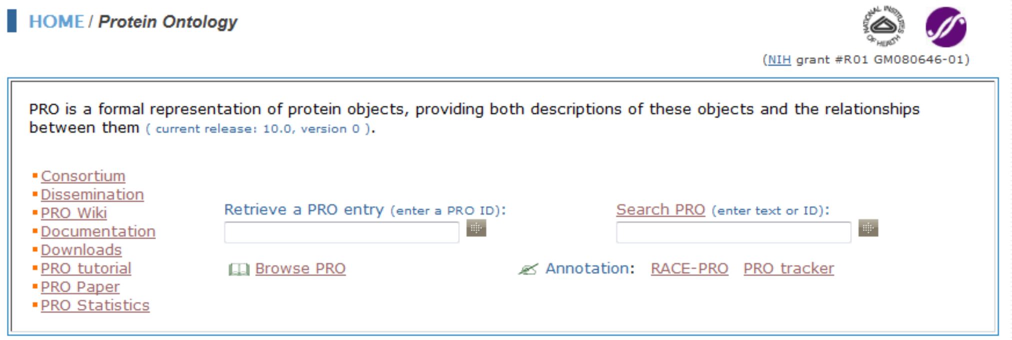
Here is an example of some of the information in this file.



PRO_ID	Object_term	Modifier	Relation	Ontology_ID	Ontology_term	Evidence_source	Evide	Taxon	DB_ID	Modified_residue, MOD_ID
PRO:000000536	c-myc isoform 1 glycosylated 1		has_function	GO:0003700	transcription factor activity	PMID:11904304	IDA	TaxID:9606	UniProtKB:P01106-1	Thr-58, MOD:00806
PRO:000000536	c-myc isoform 1 glycosylated 1		participates_in	GO:0006357	regulation of transcription from RNA polymerase II promoter	PMID:11904304	IDA	TaxID:9606	UniProtKB:P01106-1	Thr-58, MOD:00806
PRO:000000536	c-myc isoform 1 glycosylated 1		has_modification	MOD:00806	O-(N-acetylaminoglucosyl)-L-threonine	PMID:11904304		TaxID:9606	UniProtKB:P01106-1	Thr-58, MOD:00806
PRO:000000538	c-myc isoform 1 phosphorylated 2	NOT	has_function	GO:0003700	transcription factor activity	PMID:7623799	EXP	TaxID:9606	UniProtKB:P01106-1	Ser-62, MOD:00046|Thr-58, MOD:00047
PRO:000000538	c-myc isoform 1 phosphorylated 2		located_in	GO:0005634	nucleus	PMID:14563837|PMI	IDA	TaxID:10090	UniProtKB:P01108-1	Ser-62, MOD:00046|Thr-58, MOD:00047
PRO:000000538	c-myc isoform 1 phosphorylated 2		has_modification	MOD:00046	O-phospho-L-serine	PMID:14563837		TaxID:10090	UniProtKB:P01108-1	Ser-62, MOD:00046|Thr-58, MOD:00047
PRO:000000538	c-myc isoform 1 phosphorylated 2		has_modification	MOD:00047	O-phospho-L-threonine	PMID:14563837		TaxID:10090	UniProtKB:P01108-1	Ser-62, MOD:00046|Thr-58, MOD:00047

Homepage:

The menu on the left links to several documents and information pages. The link to download is there as well. The functionalities in this page are: (i) Browsing, (ii) PRO entry retrieval, (iii) text search, and (iv) annotation.



HOME / *Protein Ontology*

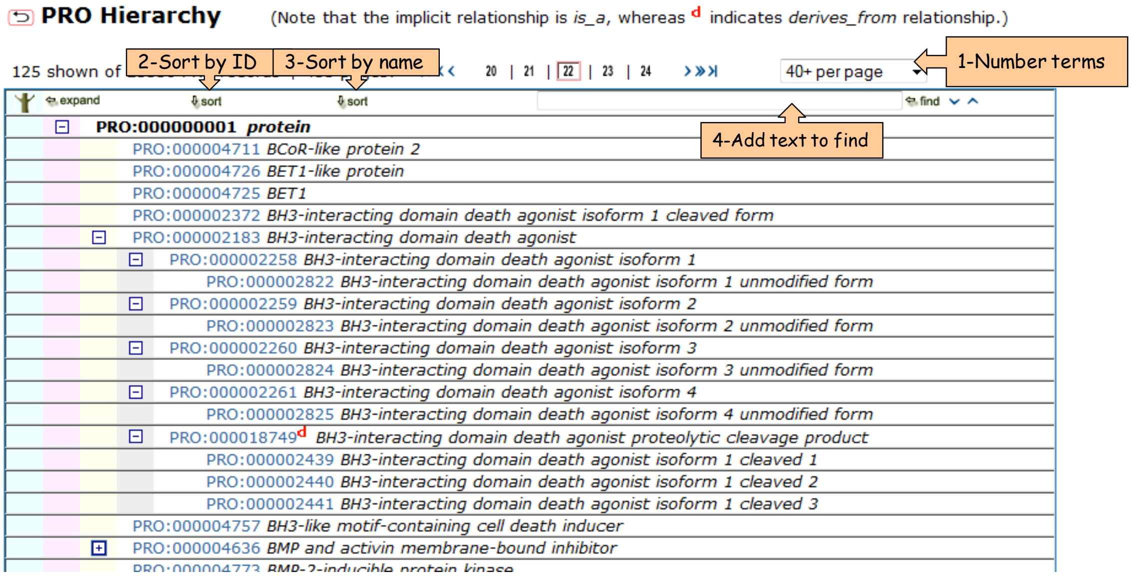
(NIH grant #R01 GM080646-01)

PRO is a formal representation of protein objects, providing both descriptions of these objects and the relationships between them (current release: 10.0, version 0).

- Consortium
- Dissemination
- PRO Wiki
- Documentation
- Downloads
- PRO tutorial
- PRO Paper
- PRO Statistics

(i) PRO Browser:

The browser is used to explore the hierarchical structure of the ontology.



(Note that the implicit relationship is *is_a*, whereas d indicates *derives_from* relationship.)

- **PRO:000000001** ***protein***
  - PRO:000004711 *BCoR-like protein 2*
  - PRO:000004726 *BET1-like protein*
  - PRO:000004725 *BET1*
  - PRO:000002372 *BH3-interacting domain death agonist isoform 1 cleaved form*
  - PRO:000002183 *BH3-interacting domain death agonist*
    - PRO:000002258 *BH3-interacting domain death agonist isoform 1*
      - PRO:000002822 *BH3-interacting domain death agonist isoform 1 unmodified form*
    - PRO:000002259 *BH3-interacting domain death agonist isoform 2*
      - PRO:000002823 *BH3-interacting domain death agonist isoform 2 unmodified form*
    - PRO:000002260 *BH3-interacting domain death agonist isoform 3*
      - PRO:000002824 *BH3-interacting domain death agonist isoform 3 unmodified form*
    - PRO:000002261 *BH3-interacting domain death agonist isoform 4*
      - PRO:000002825 *BH3-interacting domain death agonist isoform 4 unmodified form*
    - PRO:000018749d *BH3-interacting domain death agonist proteolytic cleavage product*
      - PRO:000002439 *BH3-interacting domain death agonist isoform 1 cleaved 1*
      - PRO:000002440 *BH3-interacting domain death agonist isoform 1 cleaved 2*
      - PRO:000002441 *BH3-interacting domain death agonist isoform 1 cleaved 3*
  - PRO:000004757 *BH3-like motif-containing cell death inducer*
  - PRO:000004636 *BMP and activin membrane-bound inhibitor*
  - PRO:000004773 *BMP-2-inducible protein kinase*

There are a number of functionalities that helps you to navigate the page easily:
1-Selection of number of terms per page from the drop down box on the upper right corner
2-Sorting the terms based on ID (only for the direct children of protein)
3-Sorting the terms based on term name (only for the direct children of protein)
4- Find function: it works as a text editor find option. It will highlight the exact match of the text in the various terms, by using the arrows you can go to previous pages or to next page.

The icons with a plus ⊞ and minus ⊟ signs allow expanding and collapsing nodes, respectively. Next to these icons is a PRO ID, which links to the corresponding entry report, followed by the term name. Unless otherwise stated the implicit relation between nodes is *is_a*. In the above figure **d** represents *derives_from* relationship.

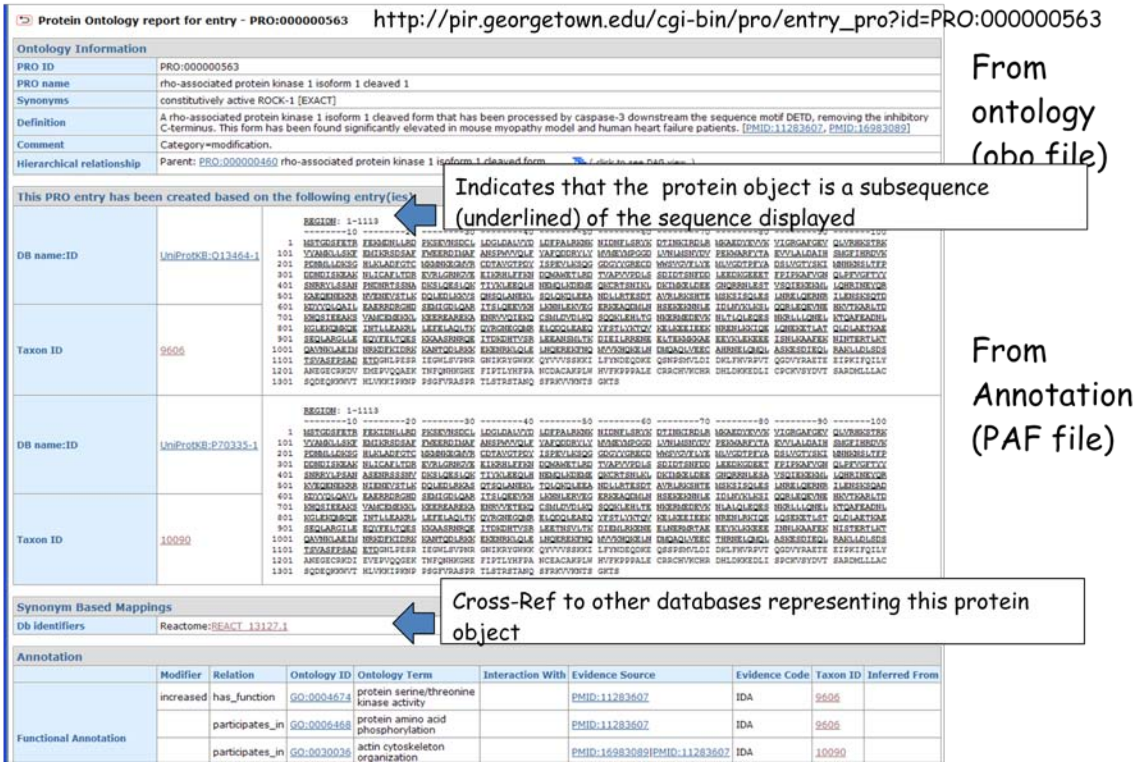
(ii) PRO entry

The PRO entry provides an integrated report about the ontology and annotation available for a given PRO term. If you know the PRO ID you can use the “retrieve PRO entry” box in the homepage. Alternatively, you can open an entry by clicking on the PRO ID in any other page (search, browser, etc). The entry report contains 4 sections:

a. Ontology information: this section displays the information from the ontology about a term (source: the pro.obo file). You can link to the parent node, to the hierarchy, and find the definition and synonyms of the term, among other things.
b. Information about the entities that were use to create the PRO entry: this section lists the sequences, in the case where category corresponds to gene, sequence or modification, for which some experimental information exists. Taxon information as well as PSI-MOD ID and modification sites are indicated when applicable. In many cases, the modifications sites are unknown and therefore only the PSI-MOD ID is listed. For cleaved products, the protein region is indicated and is underlined

in the displayed sequence. In the case of category corresponding to family, this section provides a cross-reference to the database that is the source of the class.

c. Synonymous mappings: this section contains mappings to external databases that link to protein forms as described in the given class (information source: mapping files).
d. Annotation: This section shows the annotation of the term with the different ontologies (source: PAF file). These annotations were contributed by the PRO consortium group and by community annotators through submission of RACE-PRO annotations.



Protein Ontology report for entry - PRO:000000563

http://pir.georgetown.edu/cgi-bin/pro/entry_pro?id=PRO:000000563

Ontology Information	
PRO ID	PRO:000000563
PRO name	rho-associated protein kinase 1 isoform 1 cleaved 1
Synonyms	constitutively active ROCK-1 [EXACT]
Definition	A rho-associated protein kinase 1 isoform 1 cleaved form that has been processed by caspase-3 downstream the sequence motif DETD, removing the inhibitory C-terminus. This form has been found significantly elevated in mouse myopathy model and human heart failure patients. [PMID:11283607, PMID:16983089]
Comment	Category=modification.
Hierarchical relationship	Parent: PRO:000000460 rho-associated protein kinase 1 isoform 1 cleaved form

From ontology (obo file)

This PRO entry has been created based on the following entry(ies)

DB name:ID	UniProtKB:Q13464-1
Taxon ID	9606
DB name:ID	UniProtKB:P70335-1
Taxon ID	10090

REGION: 1-1113

Indicates that the protein object is a subsequence (underlined) of the sequence displayed

From Annotation (PAF file)

Synonym Based Mappings	
Db identifiers	Reactome:REACT_13127.1

Cross-Ref to other databases representing this protein object

Annotation

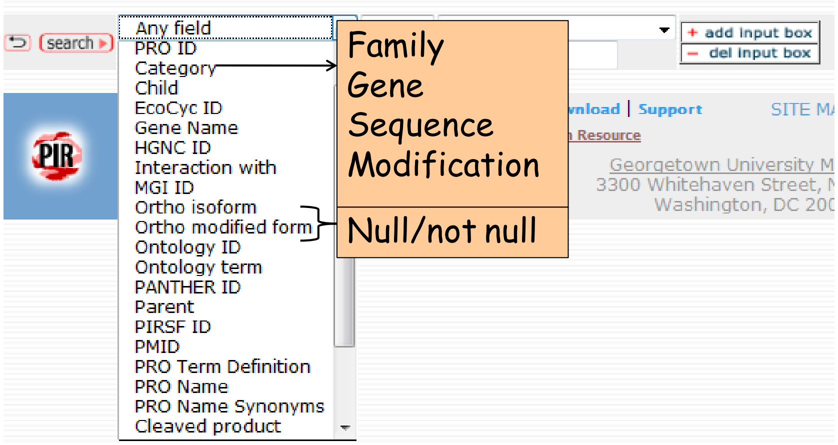
	Modifier	Relation	Ontology ID	Ontology Term	Interaction With	Evidence Source	Evidence Code	Taxon ID	Inferred From
Functional Annotation	increased	has_function	GO:0004674	protein serine/threonine kinase activity		PMID:11283607	IDA	9606	
		participates_in	GO:0006468	protein amino acid phosphorylation		PMID:11283607	IDA	9606	
		participates_in	GO:0030036	actin cytoskeleton organization		PMID:16983089|PMID:11283607	IDA	10090	

(iii) Searching PRO

The unit of search is PRO (in current setting). From the homepage you can directly enter a text in the search box and submit or you can click on the search title to access to the advance search.

The text search allows Boolean (AND, OR, NOT) searches as well as null (not present)/null (present) searches.

You have several fields available to search. Try searching some of your interest.

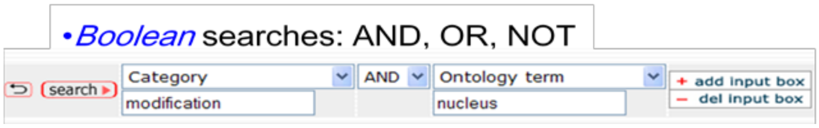


A table with explanation of each search field and examples are available at http://proconsortium.org/pro/searchPRO.pdf

AND/NOT/OR searches:

1-

Search for PRO terms that are modified forms that are annotated with GO term nucleus

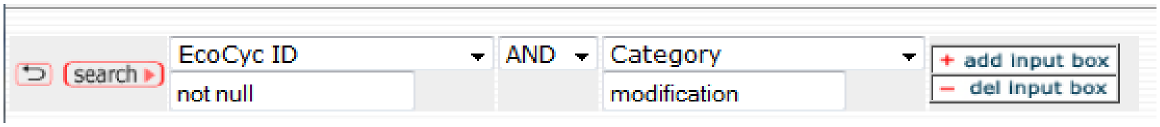


Category: modification
AND
Ontology term: nucleus

This should retrieve all PRO entries which are modified forms and contain annotation for gene ontology term nucleus.

Null/Not null searches:

- *null = absent; not null = present*



EcoCyc ID: not null
AND
Category: modification

The search above should retrieve only PRO entries that contain a reference to EcoCyc and are modified forms.

The search below should retrieve only PRO entries that have the ortho isoform tag (this information is derived from those PRO entries with isoform name, Category:sequence AND which entry contains more than one protein sequence.

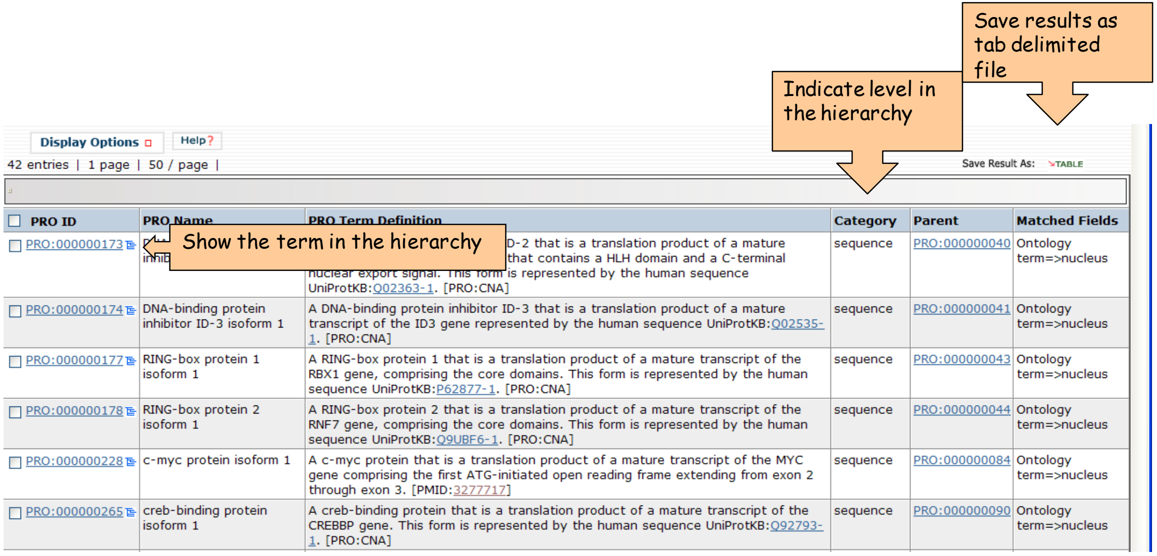
Ortho isoform: not null

AND

Any field

Result table: The default display in the result table includes the PRO ID, PRO name, PRO term definition, the category, the parent term ID and the matched field. However, this display can be changed using the display options (see below)

i. Clicking on the PRO ID opens the entry report
ii. The icon show the term in the hierarchy, i.e., opens the browser.
iii. You can save your table as a tab-delimited file using the save as table option



Save results as tab delimited file

Indicate level in the hierarchy

Show the term in the hierarchy

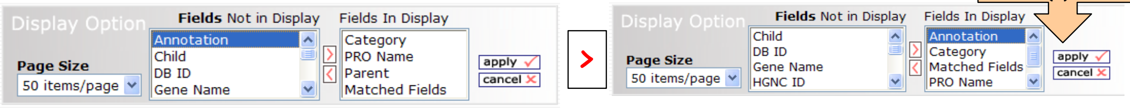
42 entries | 1 page | 50 / page |

☐ PRO ID	PRO Name	PRO Term Definition	Category	Parent	Matched Fields
☐ PRO:000000173	inhib	D-2 that is a translation product of a mature that contains a HLH domain and a C-terminal nuclear export signal. This form is represented by the human sequence UniProtKB:Q02363-1. [PRO:CNA]	sequence	PRO:000000040	Ontology term=>nucleus
☐ PRO:000000174	DNA-binding protein inhibitor ID-3 isoform 1	A DNA-binding protein inhibitor ID-3 that is a translation product of a mature transcript of the ID3 gene represented by the human sequence UniProtKB:Q02535-1. [PRO:CNA]	sequence	PRO:000000041	Ontology term=>nucleus
☐ PRO:000000177	RING-box protein 1 isoform 1	A RING-box protein 1 that is a translation product of a mature transcript of the RBX1 gene, comprising the core domains. This form is represented by the human sequence UniProtKB:P62877-1. [PRO:CNA]	sequence	PRO:000000043	Ontology term=>nucleus
☐ PRO:000000178	RING-box protein 2 isoform 1	A RING-box protein 2 that is a translation product of a mature transcript of the RNF7 gene, comprising the core domains. This form is represented by the human sequence UniProtKB:Q9UBF6-1. [PRO:CNA]	sequence	PRO:000000044	Ontology term=>nucleus
☐ PRO:000000228	c-myc protein isoform 1	A c-myc protein that is a translation product of a mature transcript of the MYC gene comprising the first ATG-initiated open reading frame extending from exon 2 through exon 3. [PMID:3277717]	sequence	PRO:000000084	Ontology term=>nucleus
☐ PRO:000000265	creb-binding protein isoform 1	A creb-binding protein that is a translation product of a mature transcript of the CREBBP gene. This form is represented by the human sequence UniProtKB:Q92793-1. [PRO:CNA]	sequence	PRO:000000090	Ontology term=>nucleus

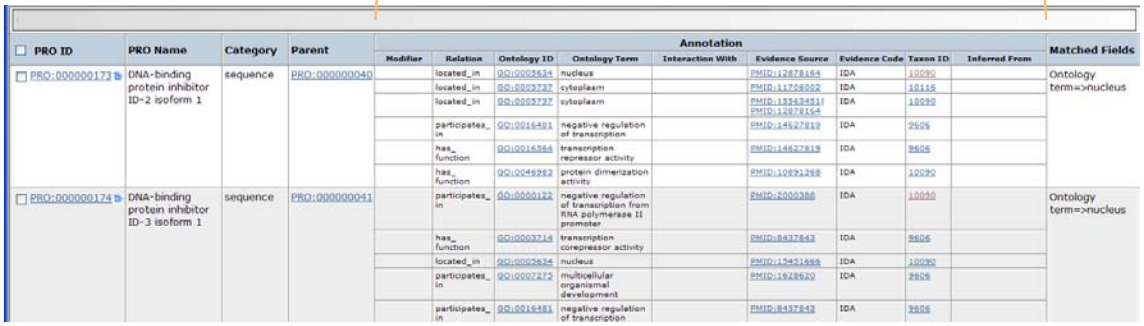
Use display options: the display options allows you to select or remove columns for the table. Use > to add or > to remove items from the list, but always select apply for the changes to take effect. In this case we have removed the definitions and the annotation has been added (only the later procedure is shown here).

Use Display Option to add/remove columns

Click apply to see the new column(s)



Annotation column added



☐ PRO ID	PRO Name	Category	Parent	Annotation									Matched Fields
				Modifier	Relation	Ontology ID	Ontology Term	Interaction With	Evidence Source	Evidence Code	Taxon ID	Inferred From	
☐ PRO:000000173	DNA-binding protein inhibitor ID-2 isoform 1	sequence	PRO:000000040		located_in	GO:0005634	nucleus		PMID:12878164	IDA	10090		Ontology term=>nucleus
					located_in	GO:0005737	cytoplasm		PMID:11706002	IDA	10116		
					located_in	GO:0005737	cytoplasm		PMID:15563451| PMID:12878164	IDA	10090		
					participates_in	GO:0016481	negative regulation of transcription		PMID:14627819	IDA	9606		
					has_function	GO:0016564	transcription repressor activity		PMID:14627819	IDA	9606		
					has_function	GO:0046983	protein dimerization activity		PMID:10891368	IDA	10090		
☐ PRO:000000174	DNA-binding protein inhibitor ID-3 isoform 1	sequence	PRO:000000041		participates_in	GO:0000122	negative regulation of transcription from RNA polymerase II promoter		PMID:2000388	IDA	10090		Ontology term=>nucleus
					has_function	GO:0003714	transcription corepressor activity		PMID:8437843	IDA	9606		
					located_in	GO:0005634	nucleus		PMID:15451666	IDA	10090		
					participates_in	GO:0007275	multicellular organismal development		PMID:1628620	IDA	9606		
					participates_in	GO:0016481	negative regulation of transcription		PMID:8437843	IDA	9606		

How to link to PRO from your database/ontology:

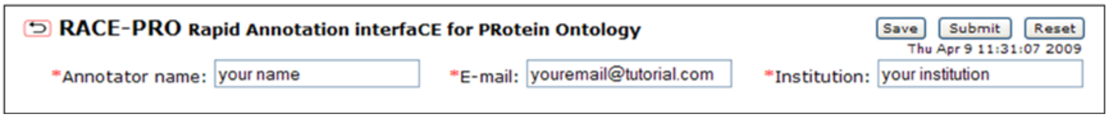
Once you have found the appropriate level to link then use the URL:
http://pir.georgetown.edu/cgi-bin/pro/entry_pro?id=PRO:xxxxxxxxx
where PRO:xxxxxxxxx is the corresponding PRO ID.
Example: http://proconsortium.org/cgi-bin/entry_pro?id=PRO:000000447

How to request or contribute to the annotation:

You can request a term or set of terms by selecting PRO tracker in the homepage. This opens the sourceforge page from OBO. You will need to provide the protein, gene name or identifier for the objects you request the terms for.
Alternatively, you can submit your request by contributing with the annotation directly. This is the purpose of the RACE-PRO page. What for?

- Obtain a PRO ID for the protein objects of interest
- Define a protein object (based on literature, experimental data)
- Add annotation on that protein object
- How it works?
    - Input your personal information (only for internal use)
    - Complete form with sequence information and annotation
    - Submit when ready (otherwise you can save for later)
    - PRO curation team will take the data, revise it, and create the corresponding PRO node in the ontology
    - Use will be informed through email about the new PRO IDs and when they will be public

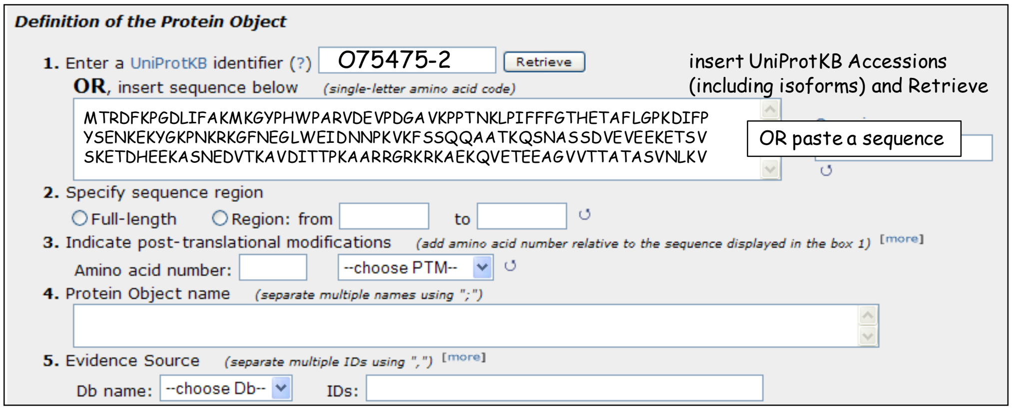
1) Fill your personal information. This allows saving your data as well as communication



Thu Apr 9 11:31:07 2009

*Annotator name: your name
*E-mail: youremail@tutorial.com
*Institution: your institution

2) Define the protein object. This allows retrieving or pasting a sequence, defining a subsequence, and/or a post-translational modification.



Definition of the Protein Object

1. Enter a UniProtKB identifier (?) O75475-2

OR, insert sequence below *(single-letter amino acid code)*

MTRDFKPGDLIFAKMKGYPHWPARVDEVPDGAVKPPTNKLPIFFFGTHETAFLGPKDIFP
YSENKEKYGKPNKRKGFNEGLWEIDNNPKVKFSSQQAATKQSNASSDVEVEEKETSV
SKETDHEEKASNEDVTKAVDITTPKAARRGRKRKAEKQVETEEAGVVTTATASVNLKV

insert UniProtKB Accessions (including isoforms) and Retrieve

OR paste a sequence

2. Specify sequence region

○ Full-length ○ Region: from to

3. Indicate post-translational modifications *(add amino acid number relative to the sequence displayed in the box 1)*

Amino acid number: --choose PTM--

4. Protein Object name *(separate multiple names using ";")*

5. Evidence Source *(separate multiple IDs using ",")*

Db name: --choose Db-- IDs:

*1.*If you use a UniProtKB identifier the sequence is formatted to show the residue numbers, and the organism box is automatically filled. You can use identifiers for isoforms as the example shown above. If you happen to have an identifier from a different database (genebank ac, etc) you can use the ID mapping or batch retrieval service from the PIR main menu (under searh/analysis). However, it is safer to paste the sequence in this case unless you know what sequence within the UniProtKB your accession corresponds to. A UniProtKB/Swiss-Prot record may contain more than one sequence which comes from different sources. Please be aware of this issue. If you use the ID mapping service you will link to the canonical sequence, but that accession could correspond to an isoform or a variant described in the file. So check the UniprotKB record.

If you paste a sequence, you need to reformat so you can see the residue numbers. Use the circle arrow to do so. Also you will need to add the organism, if you don't know the exact name you can follow the link to NCBI taxonomy browser by clicking on the Organism title.

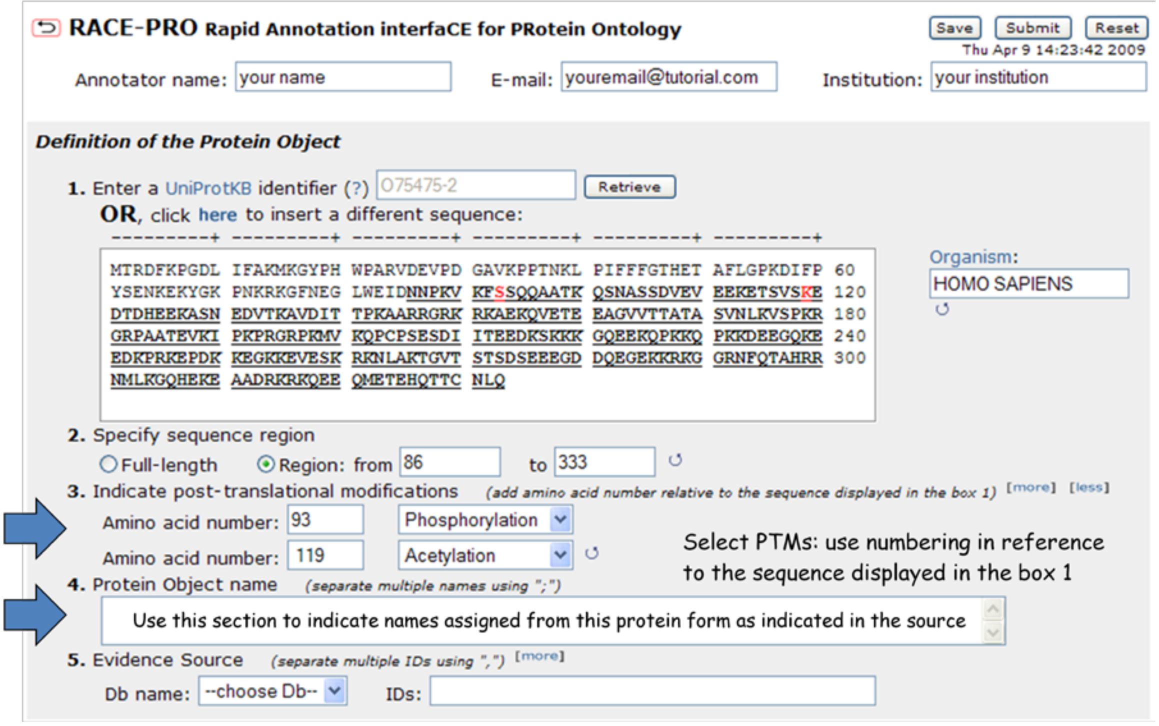
*2.*After you have the sequence displayed in the box, you can select a subsequence, for example if the protein form you are describing is a cleaved product. After you do this, click on the circle arrow to see your selected subsequence underlined.

3.Selecting the Modification: If you need to describe a modification, enter the residue number and the type of modification. If you don't know what is the residue but know about the type of modification, enter "?" in the residue number box. If the modification you need to enter is not in the list, use other to add. These terms will be later mapped to the corresponding PSI-MOD terms. For example a serine phosphorylation, will be translated into MOD:00046 phospho-L-serine.

Use the [more] to add another modification, and the [less] to remove one.

4.Protein object name: Add names by which this object is known (separated by;)

5.*DB name*: add the database source of your annotation. Select one of the options in the list, if not present use others and provide the name of the DB. In the ID box you can add many IDs separated by comma.



RACE-PRO Rapid Annotation interfaCE for PRotein Ontology

Thu Apr 9 14:23:42 2009

Annotator name: your name
E-mail: youremail@tutorial.com
Institution: your institution

Definition of the Protein Object

1. Enter a UniProtKB identifier (?) O75475-2 Retrieve

OR, click here to insert a different sequence:

```
---------+ ---------+ ---------+ ---------+ ---------+ ---------+
MTRDFKPGDL IFAKMKGYPH WPARVDEVPD GAVKPPTNKL PIFFFGTHET AFLGPKDIFP 60
YSENKEKYGK PNKRKGFNEG LWEIDNNPKV KFSSQQAATK QSNASSDVEV EEKETSVSKE 120
DTDHEEKASN EDVTKAVDIT TPKAARRGRK RKAEKQVETE EAGVVTTATA SVNLKVSPKR 180
GRPAATEVKI PKPRGRPKMV KQPCPSESDI ITEEDKSKKK GQEEKQPKKQ PKKDEEGQKE 240
EDKPRKEPDK KEGKKEVESK RKNLAKTGVT STSDSEEEGD DQEGEKKRKG GRNFQTAHRR 300
NMLKGQHEKE AADRKRKQEE QMETEHQTTC NLQ
```

Organism: HOMO SAPIENS

2. Specify sequence region

○ Full-length ● Region: from 86 to 333

3. Indicate post-translational modifications *(add amino acid number relative to the sequence displayed in the box 1)* [more] [less]

Amino acid number: 93 Phosphorylation

Amino acid number: 119 Acetylation

Select PTMs: use numbering in reference to the sequence displayed in the box 1

4. Protein Object name *(separate multiple names using ";")*

Use this section to indicate names assigned from this protein form as indicated in the source

5. Evidence Source *(separate multiple IDs using ",")* [more]

Db name: --choose Db-- IDs:

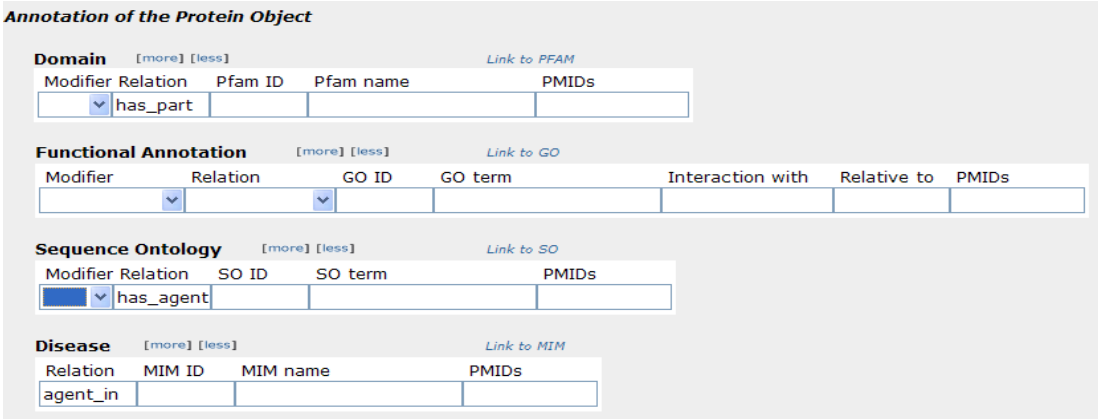
The second block is the annotation.

The annotations are separated by database/ontology: domain, functional terms, sequence and disease. You only add those that are pertinent to the information you have from the source.

All the information about the different fields is in the PAFguidelines. But below are some of clarifications:

- Modifiers: used to modify a relation between a PRO term and another term. It includes the GO qualifiers NOT, contributes_to plus increased, decreased, and altered (to be used with the relative to column).
- Relation to the specific annotation. For some database/ontology there is a single relation and that is displayed
- Add ID for the specific database/ontology. If you need to search use the "link to .." link. Future development: autofill of name.

Use the [more], [less] to add or remove an annotation line.



Annotation of the Protein Object

Domain [more] [less] Link to PFAM

Modifier	Relation	Pfam ID	Pfam name	PMIDs
	has_part			

Functional Annotation [more] [less] Link to GO

Modifier	Relation	GO ID	GO term	Interaction with	Relative to	PMIDs

Sequence Ontology [more] [less] Link to SO

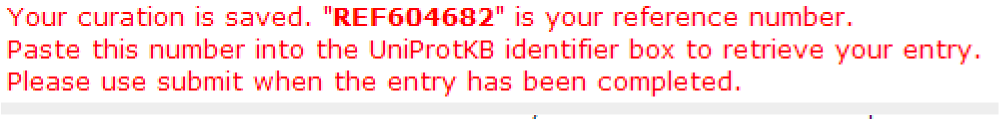
Modifier	Relation	SO ID	SO term	PMIDs
	has_agent			

Disease [more] [less] Link to MIM

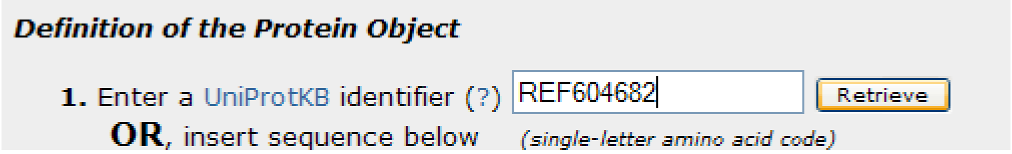
Relation	MIM ID	MIM name	PMIDs
agent_in			

5.Saving / submitting the annotation:

Save is to give you the possibility to save your data in case you have not finished and need to return to the annotation later. When you save you are given a REF number and then you can insert this number in the UniProtKB identifier box to retrieve your entry.



Your curation is saved. "**REF604682**" is your reference number.
Paste this number into the UniProtKB identifier box to retrieve your entry.
Please use submit when the entry has been completed.



Definition of the Protein Object

1. Enter a UniProtKB identifier (?) REF604682

OR, insert sequence below *(single-letter amino acid code)*

Submit is used when the user is done.

What happens next?

Receive an email with the ref number in the subject when your entry is under reviewed

A PRO curator will be assigned to review your entry and create the corresponding PRO node.

You will receive an email with the PRO ID, and the terms for your final check

Now try creating a RACE-PRO entry for one of the forms of the human serase-1b, a splice variant of the TMPRSS9 gene based on the paper corresponding to PMID:16872279. Link to the paper http://www.biochemj.org/bj/400/0551/bj4000551.htm

Done!

Note:Thanks to Judy Blake for providing a template for the tutorial.