Welcome to PRO Tutorial

Cecilia Arighi, Ph.D
Head of PRO curation team
arighi@dbi.udel.edu

Outline

Introduction to the PRotein Ontology
PRO Website
  PRO Entry
  Search and Browse
Annotation
  Annotation tool: RACE-PRO
Annotation group exercise

All documents available at:
http://pir.georgetown.edu/~arighic/pro/tutorial/
### PRO within OBO Foundry

Ontology for semantic integration of heterogeneous biological data

OBO Foundry establishes rules and best practices to create a suite of orthogonal interoperable reference ontologies

<table>
<thead>
<tr>
<th>RELATION TO TIME</th>
<th>CONTINUANT</th>
<th>OCURRENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRANULARITY</td>
<td>INDEPENDENT</td>
<td>DEPENDENT</td>
</tr>
<tr>
<td>ORGAN AND ORGANISM</td>
<td>Organism (NCBI Taxonomy)</td>
<td>Anatomical Entity (FMA, CARO)</td>
</tr>
<tr>
<td>CELL AND CELLULAR COMPONENT</td>
<td>Cell (CL)</td>
<td>Cellular Component (FMA, GO)</td>
</tr>
<tr>
<td>MOLECULE</td>
<td>Molecule (ChEBI, SO, RNAO, PRO)</td>
<td>Molecular Function (GO)</td>
</tr>
</tbody>
</table>

**PRotein Ontology (PRO)**

**PRO** in OBO Foundry to represent **protein entities**

**Ontology for Protein Evolution (ProEvo)** for **evolutionary classes of proteins**: captures the **protein classes reflecting** evolutionary relationships at full-length protein levels. **It is meant to indicate the relatedness of proteins, not their evolution.** Use **is_a** relationship.

**Ontology for Protein Forms (ProForm)** for **multiple protein forms of a gene**: captures the different protein forms of a specific gene from genetic variations, alternative splicing, cleavage, post-translational modifications. Relations used **is_a** or **derives_from**

Why PRO?

Allow specification of relationships between PRO and other ontologies, such as GO and Disease Ontology

Provides a structure to support formal, computer-based inferences based on shared attributes among homologous proteins

Provides a stable unique identifier to any protein type

PRO:000000650 Smad2 isoform 1 phosphorylated 1 (phosphorylated SSxS motif) NOT has_function GO:0003677 DNA binding

PRO:000000656 Smad2 isoform 2 phosphorylated 1 (phosphorylated SSxS motif) has_function GO:0003677 DNA binding

Provides formalization and precise annotation of specific protein forms/classes, allowing accurate and consistent data mapping, integration and analysis: for dendritic cell ontology or pathway

http://www.biomedcentral.com/1471-2105/10/70

TGF-beta signaling – comparison between PID and Reactome

Growth signals

Ca2+

MEKK1

ERK1/2

CaM

TGF-beta receptor

Smad 2

Smad 4

Growth signals

Stress signals

Cytoplasm

Nucleus

DNA binding and transcription regulation

MAPKKK

P38 MAPK pathway

JNK cascade

MAPKKK

XIAP

TAK1

Degradation

Phosphorylation (P) at Serine (S), Threonine (T), Tyrosine (Y)
Ubiquitination (U) at Lysine (K)

Common in both Reactome & PID
Only included in Reactome

* All others are in PID. Not all components in the pathway from both databases are listed.
The Ontology Structure

protein

is_a protein family level (category=family)

is_a protein subfamily level

is_a gene product level (category=gene)

isoform 1 (category=sequence)

derives_from isoform 1 PTM x form (category=modification)

is_a isoform 1 PTM x 1

is_a isoform 1 PTM x 2

is_a isoform 1 PTM y form

is_a isoform 1 PTM y 1

is_a isoform 2 (category=sequence)

is_a sequence variant 1 (category=sequence)

Read bottom-up

Future introduction of species-specific level
Some definitions

**Categories:** indication of how PRO is organized

*Family:* a PRO term at this level refers to proteins that can trace back to a common ancestor over the entire length of the protein, are part of the same family.

*Gene:* a PRO term at this level refers to the protein products of a distinct gene.

*Sequence:* a PRO term at this level refers to the protein products with a distinct sequence upon initial translation.

*Modification:* a PRO term at this level refers to the protein products with some change that occurs after initial translation.

**Ortho-isoform:** These are isoforms encoded by orthologous genes that are believed to have arisen prior to speciation and divergence of the primary sequence.

**Ortho-modified form:** Post-translational modifications on equivalent residues in ortho-isoforms.
Annotation of PRO through other ontologies

**PRO**

**Root Level**

**Family-Level Distinction**
- Derivation: common ancestor
- Source: PIRSF family

**Gene-Level Distinction**
- Derivation: specific gene
- Sources: PIRSF subfamily, Panther subfamily

**Sequence-Level Distinction**
- Derivation: specific allele or splice variant
- Source: UniProtKB

**Modification-Level Distinction**
- Derived from post-translational modification
- Source: UniProtKB

**PAF txt**

- **Pfam**
  - Domain
  - protein domain
  - has_part

- **GO**
  - Gene Ontology
  - molecular function
  - biological process
  - cellular component
  - participates_in
  - located_in

- **OMIM**
  - Disease
  - disease
  - agent_in

- **SO**
  - Sequence Ontology
  - sequence change
  - has_agent (sequence change)

- **PSI-MOD**
  - Modification
  - protein modification
  - has_modification
PRO distribution files
ftp://ftp.pir.georgetown.edu/databases/ontology/proobo/

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
### PRO distribution files

**PAF.txt** → **PRO association file (Tab delimited, similar to GAF)**

<table>
<thead>
<tr>
<th>Column</th>
<th>Column Title</th>
<th>Column</th>
<th>Column Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PRO_ID</td>
<td>12</td>
<td>Inferred_from</td>
</tr>
<tr>
<td>2</td>
<td>Object_term</td>
<td>13</td>
<td>DB_ID</td>
</tr>
<tr>
<td>3</td>
<td>Object_syny</td>
<td>14</td>
<td>Protein_region</td>
</tr>
<tr>
<td>4</td>
<td>Modifier</td>
<td>15</td>
<td>Modified_residue, MOD_ID</td>
</tr>
<tr>
<td>5</td>
<td>Relation</td>
<td>16</td>
<td>Date</td>
</tr>
<tr>
<td>6</td>
<td>Ontology_ID</td>
<td>17</td>
<td>Assigned_by</td>
</tr>
<tr>
<td>7</td>
<td>Ontology_term</td>
<td>18</td>
<td>Equivalent_Form</td>
</tr>
<tr>
<td>8</td>
<td>Interaction_with</td>
<td>19</td>
<td>Comment</td>
</tr>
<tr>
<td>9</td>
<td>Evidence_source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Evidence_code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Taxon</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

**Current scope:** human, mouse and E. coli proteins. It will have comprehensive coverage for gene and isoform level.

**Prototype:** TGF-beta signaling pathway to design the framework

79 proteins retrieved from KEGG (include BMP, TGF beta and activin pathways).

**Current release 5.0 contain 2136 terms with their respective definitions, comprising the families and translation products for 980.**

<table>
<thead>
<tr>
<th>Category</th>
<th>Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>206</td>
</tr>
<tr>
<td>Gene</td>
<td>980</td>
</tr>
<tr>
<td>Sequence</td>
<td>616</td>
</tr>
<tr>
<td>Modification</td>
<td>334</td>
</tr>
</tbody>
</table>
What is Annotated?

- Pfam for ProEvo and ProForm
- GO terms with ISS if ProEvo
- GO terms with EXP evidence if ProForm
- PSI-MOD terms for protein modifications
- SO for sequence variants
- MIM for sequence variants

No ontologies available

Annotate this terms

MOD Ontology

<table>
<thead>
<tr>
<th>MOD ID</th>
<th>MOD Term</th>
<th>Qualifier</th>
<th>Relation</th>
<th>PMID</th>
<th>Taxon ID</th>
<th>Evidence Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD:00046</td>
<td>O-phospho-L-serine</td>
<td>-</td>
<td>has_modification</td>
<td>9873005</td>
<td>9606</td>
<td>TAS</td>
</tr>
</tbody>
</table>

Domain

GO Ontology

<table>
<thead>
<tr>
<th>GO ID</th>
<th>GO Term</th>
<th>Qualifier</th>
<th>Relation</th>
<th>PMID</th>
<th>Taxon ID</th>
<th>Evidence Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0003677</td>
<td>DNA binding</td>
<td>-</td>
<td>has_function</td>
<td>14701940,9873005</td>
<td>9606</td>
<td>EXP</td>
</tr>
</tbody>
</table>

Sequence Ontology
SONI-MOD

Ontology that describes the possible modifications to protein amino acid residues

Use PTM list in UnIProtKB to find MOD term

SO

Ontology that can describe the possible causes of protein sequence, expression, or structure changes
Curation workflow

PrePRO, automatic generated file

1) Automatic procedure to extract information from:
- UniProt
- Source for ProForm terms
- Sequence variants and splice forms
- PTMs, GO and Disease annotation
- PIRSF
- Source for ProEvo terms
- PIRSF-Pfam mapping
- MGI
- Source of orthology
- GO annotation to isoforms

2) Manual curation:
- Use OBO editor 1.1
- Check, delete and/or create terms
- Create definitions: ProEvo terms are based on domain composition,
  ProForm terms are based on presence/absence of protein features.
- Assign the existing annotation to the specific PRO node
- Add additional annotation/forms based on literature search
Protein ontology web-based editor
(almost there)

http://pir.georgetown.edu/pro/editor.shtml
Protein Ontology Community Annotation

Retrieve PRO Terms for Curation

1. Enter IDs: (maximum 100)

Or an ID file: Browse...

2. Specify the ID type:
   - UniProtKB ACC
   - PRO ID

Submit  Reset
Add, remove GO terms using pencil.
### Curation Form for PrePRO:900120603

<table>
<thead>
<tr>
<th>name</th>
<th>mothers against decapentaplegic homolog 2</th>
</tr>
</thead>
</table>
| synonym | "MAD homolog 2" EXACT []
"Mad-related protein 2" EXACT []
"SMAD 2" EXACT []
"mMad2" EXACT []
"Mothers against DPP homolog 2" EXACT []
"Mothers against decapentaplegic homolog 2" EXACT []
"Smad2" RELATED[]
"Madh2, Madr2" RELATED[] |
| namespace | pro |
| def | "A receptor-regulated Smad protein, Smad 2/Smad 3 types that is a translation product of the SMAD2 gene." [PRO:PrePRO] |
| is_a | PrePRO:900019297 |

**Db identifiers:** HGNC:6768; MGI:108051; UniProtKB:Q62432

**Related PMID:**

#### Annotate this term
- Domain [Link to PFAM]
- Gene Ontology [Link to GO]
- Sequence Ontology [Link to SO]

**Parent:** PrePRO:900019297

**Children:** PrePRO:900120605
PrePRO:900120604

**Link to databases. You can add new X-ref**
What is curated in the ontology?

**Name:** the data source name. For upper ProEvo classes we use PIRSF and/or PANTHER names, for gene product level we use UniprotKB/Swiss-prot. Names start in lowercase, except for ATP, DNA, etc.

**Synonyms:** imported from UniprotKB for the gene level

**Definitions:** we try to create standard definitions. Some of them are provided as templates.

\[ A \text{ is a } B \text{ that } C \text{'s} \]

**Cross-ref:** to PIRSF/PANTHER for family and gene level

*But First curation of the ontological structure is needed:

- Merging of ProEvo nodes
- Merging of ProForm nodes: need to find equivalent isoforms (ortho-isoforms)
- Define and add modified form nodes: what isoform and what modification
**Define and add modified form nodes**

**Feature section of UniProtKB record for human smad2**

<table>
<thead>
<tr>
<th>Amino acid modifications</th>
<th>Modified residue</th>
<th>2</th>
<th>1</th>
<th>N-acetylserine [Ref.7]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified residue</td>
<td>8</td>
<td>1</td>
<td>Phosphothreonine; by MAPK3 [Ref.14] [Ref.17] [Ref.20]</td>
<td></td>
</tr>
<tr>
<td>Modified residue</td>
<td>19</td>
<td>1</td>
<td>N\textsubscript{6}-acetyllysine [Ref.18]</td>
<td></td>
</tr>
<tr>
<td>Modified residue</td>
<td>220</td>
<td>1</td>
<td>Phosphothreonine [Probable]</td>
<td></td>
</tr>
<tr>
<td>Modified residue</td>
<td>240</td>
<td>1</td>
<td>Phosphoserine; by CAMK2 [Ref.13]</td>
<td></td>
</tr>
<tr>
<td>Modified residue</td>
<td>245</td>
<td>1</td>
<td>Phosphoserine [Probable]</td>
<td></td>
</tr>
<tr>
<td>Modified residue</td>
<td>250</td>
<td>1</td>
<td>Phosphoserine [Probable]</td>
<td></td>
</tr>
<tr>
<td>Modified residue</td>
<td>255</td>
<td>1</td>
<td>Phosphoserine [Probable]</td>
<td></td>
</tr>
<tr>
<td>Modified residue</td>
<td>458</td>
<td>1</td>
<td>Phosphoserine [Ref.20] [Ref.21]</td>
<td></td>
</tr>
<tr>
<td>Modified residue</td>
<td>465</td>
<td>1</td>
<td>Phosphoserine [Ref.9] [Ref.10]</td>
<td></td>
</tr>
<tr>
<td>Modified residue</td>
<td>467</td>
<td>1</td>
<td>Phosphoserine [Ref.9] [Ref.10]</td>
<td></td>
</tr>
</tbody>
</table>

**Reference section**

\textit{Cited for: PHOSPHORYLATION AT SER-465 AND SER-467.}"

\textit{Cited for: PHOSPHORYLATION AT SER-240.}"
ProForm Curation

by TGF-beta receptor through Ca++-mediated signaling
PRO homepage

PRO is a formal representation of protein objects, providing both descriptions of these objects and the relationships between them.

- Consortium
- Dissemination
- PRO Wiki
- Documentation
- Downloads

Retrieve a PRO entry (enter a PRO ID):

Search PRO (enter text or ID):

Annotation: RACE-PRO PRO tracker

PRO encompasses a sub-ontology of proteins based on evolutionary relatedness (ProEvo) and a sub-ontology of the multiple protein forms produced from a given gene (ProForm). PRO is interoperable with other OBO Foundry ontologies such as the Sequence Ontology (SO) and the Gene Ontology (GO) that provide representations of protein qualities. This interoperability facilitates cross-species comparisons, pathway analysis, disease modeling, and the generation of new hypotheses through data integration and machine reasoning.
## PRO Entry

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

### Ontology Information

<table>
<thead>
<tr>
<th>PRO ID</th>
<th>Name</th>
<th>Synonyms</th>
<th>Definition</th>
<th>Annotation</th>
<th>CrossRef</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRO:000000563</td>
<td>rho-associated protein kinase 1 isoform 1 deaved 1</td>
<td>constitutively active ROCK-1 [EXACT]</td>
<td>A rho-associated protein kinase 1 isoform 1 deaved 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 myoathym model and human heart failure patients. [PMID:11283607, PMID:10983389]</td>
<td>Indicates that the protein object is a subsequence (underlined) of the sequence displayed.</td>
<td>CrossRef to other databases representing this protein object.</td>
</tr>
</tbody>
</table>

### Annotation

<table>
<thead>
<tr>
<th>Modifier</th>
<th>Relation</th>
<th>Ontology ID</th>
<th>Ontology Term</th>
<th>Interaction With</th>
<th>Evidence Source</th>
<th>Evidence Code</th>
<th>Taxon ID</th>
<th>Inferred From</th>
</tr>
</thead>
<tbody>
<tr>
<td>increased</td>
<td>has_function</td>
<td>GO:0004674</td>
<td>protein serine/threonine kinase activity</td>
<td>PMID:11283607</td>
<td>IDA</td>
<td>9606</td>
<td></td>
<td></td>
</tr>
<tr>
<td>participates_in</td>
<td>has_function</td>
<td>GO:0004648</td>
<td>protein amino acid phosphorylation</td>
<td>PMID:11283607</td>
<td>IDA</td>
<td>9606</td>
<td></td>
<td></td>
</tr>
<tr>
<td>participates_in</td>
<td>has_function</td>
<td>GO:00036035</td>
<td>actin cytoskeleton organization</td>
<td>PMID:16983089, PMID:11283607</td>
<td>IDA</td>
<td>10090</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PRO homepage

PRO is a formal representation of protein objects, providing both descriptions of these objects and the relationships between them.

- Consortium
- Dissemination
- PRO Wiki
- Documentation
- Downloads
- Browse PRO
- Annotation: RACE-PRO PRO tracker

PRO encompasses a sub-ontology of proteins based on evolutionary relatedness (ProEvo) and a sub-ontology of the multiple protein forms produced from a given gene (ProForm). PRO is interoperable with other OBO Foundry ontologies--such as the Sequence Ontology (SO) and the Gene Ontology (GO)--that provide representations of protein qualities. This interoperability facilitates cross-species comparisons, pathway analysis, disease modeling, and the generation of new hypotheses through data integration and machine reasoning.
Search for PRO terms that are modified forms that are annotated with GO term nucleus

http://pir.georgetown.edu/cgi-bin/pro/search_pro

• **Boolean** searches: AND, OR, NOT
**Use Display Option to add/remove columns**

<table>
<thead>
<tr>
<th>PRO ID</th>
<th>PRO Name</th>
<th>PRO Term Definition</th>
<th>Category</th>
<th>Parent</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRO:00000507</td>
<td>DNA-binding protein inhibitor ID-2 isoform 1 phosphorylated 1</td>
<td>A DNA-binding protein inhibitor ID-2 isoform 1 phosphorylated form that has been phosphorylated at the Ser residue in the N-terminal SPVR motif. [PMID:9029153]</td>
<td>modification</td>
<td>PRO:000000392</td>
<td>located_in, GO:0005624, nucleus, PMID:11706002, IDA</td>
</tr>
<tr>
<td>PRO:00000511</td>
<td>RING-box protein 2 isoform 1 phosphorylated 1</td>
<td>A RING-box protein 2 isoform 1 phosphorylated form that has been phosphorylated at the most N-terminal Thr residue by Ck2. [PMID:12748192]</td>
<td>modification</td>
<td>PRO:000000396</td>
<td>located_in, GO:0005634, nucleus, PMID:12748192, IDA</td>
</tr>
</tbody>
</table>

**Annotation column added**

Click apply to see the new column(s)
Search for PRO terms with mapping to Reactome

- \textit{null} = absent; \textit{nonnull} = present

<table>
<thead>
<tr>
<th>PRO ID</th>
<th>PRO Name</th>
<th>Category</th>
<th>Parent</th>
<th>Reactome ID</th>
<th>Matched Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRO:000000099</td>
<td>inhibitory smad protein</td>
<td>family</td>
<td>PRO:00000027</td>
<td>REACT:7835.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000125</td>
<td>sara</td>
<td>gene</td>
<td>PRO:00000026</td>
<td>REACT:7209.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000127</td>
<td>smad ubiquitination regulatory factor</td>
<td>family</td>
<td>PRO:00000016</td>
<td>REACT:7705.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000136</td>
<td>BMP receptor type-1A isoform 1</td>
<td>sequence</td>
<td>PRO:00000035</td>
<td>REACT:12185.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000148</td>
<td>BMP receptor type-2 isoform 1</td>
<td>sequence</td>
<td>PRO:00000032</td>
<td>REACT:12286.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000177</td>
<td>RING-box protein 1 isoform 1</td>
<td>sequence</td>
<td>PRO:00000043</td>
<td>REACT:9719.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000180</td>
<td>5-phase kinase-associated protein 1A isoform 1</td>
<td>sequence</td>
<td>PRO:00000045</td>
<td>REACT:7311.1; REACT:9096.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000228</td>
<td>c-myc protein isoform 1</td>
<td>sequence</td>
<td>PRO:00000064</td>
<td>REACT:9224.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000247</td>
<td>cullin 1 isoform 1</td>
<td>sequence</td>
<td>PRO:00000092</td>
<td>REACT:7558.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000268</td>
<td>cyclin-dependent kinase 4 inhibitor 8 isoform 1</td>
<td>sequence</td>
<td>PRO:00000093</td>
<td>REACT:8570.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000289</td>
<td>mitogen-activated protein kinase 1 isoform 1</td>
<td>sequence</td>
<td>PRO:00000103</td>
<td>REACT:7897.1; REACT:5485.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000290</td>
<td>mitogen-activated protein kinase 3 isoform 1</td>
<td>sequence</td>
<td>PRO:00000104</td>
<td>REACT:4130.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000307</td>
<td>rho-associated protein kinase 1 isoform 1</td>
<td>sequence</td>
<td>PRO:00000120</td>
<td>REACT:13335.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000310</td>
<td>ribosomal protein S6 kinase beta 1 isoform 1</td>
<td>sequence</td>
<td>PRO:00000123</td>
<td>REACT:7863.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000312</td>
<td>ribosomal protein S6 kinase beta 2 isoform 1</td>
<td>sequence</td>
<td>PRO:00000124</td>
<td>REACT:13292.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000313</td>
<td>sara isoform 1</td>
<td>sequence</td>
<td>PRO:00000125</td>
<td>REACT:7209.1; REACT:7369.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000314</td>
<td>sara isoform 2</td>
<td>sequence</td>
<td>PRO:00000125</td>
<td>REACT:7371.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
<tr>
<td>PRO:000000315</td>
<td>sara isoform 3</td>
<td>sequence</td>
<td>PRO:00000125</td>
<td>REACT:6985.1</td>
<td>Reactome ID=&gt;null</td>
</tr>
</tbody>
</table>
Search for PRO terms with ortho isoforms

### Protein Ontology report for entry - PRO:000000048

**Ontology Information**
- **PRO ID**: PRO:000000048
- **PRO name**: TGF-beta receptor type-2 isoform 1
- **Synonyms**: Isoform RII-1 [EXACT]
- **Definition**: A TGF-beta receptor type-2 that is a translation product of a mature transcript of the TGFBR2 gene, and that includes the core domains. This form is represented by the mouse sequence UniProtKB:Q62312-2. [PMID:7957954]
- **Comment**: Category=sequence.

**Hierarchical relationship**: Parent: PRO:000000048 TGF-beta receptor-type 2

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

<table>
<thead>
<tr>
<th>DB name:ID</th>
<th>UniProtKB:O23312-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxon ID</td>
<td>10000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DB name:ID</th>
<th>UniProtKB:P37173-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxon ID</td>
<td>9606</td>
</tr>
</tbody>
</table>

**Synonym Based Mappings**
- **Db identifiers**: UniProtKB:P37173-1, Q62312-2

**Annotation**

<table>
<thead>
<tr>
<th>Functional Annotation</th>
<th>Modifier</th>
<th>Relation</th>
<th>Ontology ID</th>
<th>Ontology Term</th>
<th>Interaction With</th>
<th>Evidence Source</th>
<th>Evidence Code</th>
<th>Taxon ID</th>
<th>Inferred From</th>
</tr>
</thead>
<tbody>
<tr>
<td>located_in</td>
<td></td>
<td>integral to membrane</td>
<td>GO:0016021</td>
<td></td>
<td>PMID:7957954</td>
<td>EXP</td>
<td>10000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PRO homepage

PRO is a formal representation of protein objects, providing both hierarchical organization and relationships between them.

- Consortium
- Dissemination
- PRO Wiki
- Documentation
- Downloads

Retrieve a PRO entry (enter a PRO ID):
Search PRO (enter text or ID):

Annotation tool

PRO encompasses a sub-ontology of proteins based on evolutionary relatedness (ProEvo) and a sub-ontology of the multiple protein forms produced from a given gene (ProForm). PRO is interoperable with other OBO Foundry ontologies such as the Sequence Ontology (SO) and the Gene Ontology (GO)—that provide representations of protein qualities. This interoperability facilitates cross-species comparisons, pathway analysis, disease modeling, and the generation of new hypotheses through data integration and machine reasoning.

PRO

Protein Ontology

Root Level

Family-Level Distinction
- Derivation: common ancestor
  - Source: PFAM family

Gene-Level Distinction
- Derivation: specific gene
  - Source: PFAM subfamily, Panther subfamily

Sequence-Level Distinction
- Derivation: specific splice variant
  - Source: UniProtKB

Modification-Level Distinction
- Derived from post-translational modification
  - Source: UniProtKB

ProForm

Modification Level

Sequence Level

Gene Level

ProEvo

Family Level

Root Level

Example:
- TGF-beta receptor phosphorylated smad2 isoform1
  - phosphorylated
  - smad2
  - isoform1

Pfam

protein domain

GO

molecular function

biological process

MIM

disease

SO

sequence change

PSI-MOD

protein modification
Annotation tool for PRO: RACE-PRO

Obtain a PRO ID for the protein objects of interest
Define a protein object (based on literature, experimental data)
Add annotation to that protein object

How does it work?

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 ID and when they will be public
Annotation tool for PRO: RACE-PRO

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

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


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) [more]
   - Amino acid number: [ ] --choose PTM--

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

5. Evidence Source (separate multiple IDs using ",") [more]
   - Db name: --choose Db--
   - IDs: [ ]
Annotation tool for PRO: RACE-PRO

RACE-PRO Rapid Annotation interface for PRotein Ontology

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

Definition of the Protein Object

O75475-2 PSIP1_HUMAN ISOFORM 2 OF PC4 AND SFRS1-INTERACTING PROTEIN OS value="">

1. Enter a UniProtKB identifier (?) OR, insert sequence below (single-letter amino acid code)

```
MTRDFKPGDL IFAKMKGYPH WPARVDEVPD GAVKPPTNKL PIFFGTHET AFLGPKDIFP 60
YSENKEYGK PKRKKGFNEG LWEIDNNPKV KFPSSQAATK QSNAADDVEV EERETSVSKE 120
DTDHEEKSASN EDVTKAVDIT TPKAARRGRK RKAEBKVETE HAGVVTTATA SVNLKVSFKR 180
GRPAATEVTK PKPRGRPMV KQPCPSESDI ITEEDKSKKK GQEBKQPKKQ PKRDEEGQKE 240
```

Organism: HOMO SAPIENS

2. Specify sequence region

- Full-length  - Region: from 86 to 333 Select a region (e.g. if a cleaved product)

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 ";"): [more]

Db name: --choose Db-- IDs:
Annotation tool for PRO: RACE-PRO

RACE-PRO Rapid Annotation interface for PRotein Ontology

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

Definition of the Protein Object

1. Enter a UniProtKB identifier (?) 075475-2
   OR, click here to insert a different sequence:
   MTRDFKPQDL IFAKMGYPHP WAPARVDEVPL GAVKPPTNKL P1F5GTHET AFLGPDKDFF 60
   YSNKEKYKPG PNKRKGFNEG LWEIDNNPKV KF5Q0AAPK QSNASHDDV EVKBTSVSKR 120
   DTDHEKASN EDVTKAVDIT TPQKARRGRK RKAEOQVTH EAGTVTAT A2NLKQSPKR 180
   GRPAATEVM PKQPRGRPKMV KQPQCESDIT ITEEKSSKQ GQEEKDPKKQ PKDEEQQKEK 240
   EDKPRREPDPK KEKKKEFESK RKNLAKGVT STSDSEEGD QDEGEKRRKQ GRNFQATHRR 300
   NMLKQHEHEK ADRKRRQFQF QMEHTQHTC NLO

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)
   Amino acid number: 93
   Phosphorylation √
   Amino acid number: 119
   Acetylation √

4. Protein Object name
   (separate multiple names using “;”)

5. Evidence Source
   (separate multiple IDs using “,”)
   Db name: --choose Db--
   IDs:

Non real example

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

Select PTM use numbering in reference to the sequence displayed in the box 1
**Annotation tool for PRO: RACE-PRO**

**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.

### Annotation of the Protein Object

<table>
<thead>
<tr>
<th>Domain</th>
<th>[more] [less]</th>
<th>Link to PFAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modifier Relation</td>
<td>Pfam ID</td>
<td>Pfam name</td>
</tr>
<tr>
<td>has_part</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Functional Annotation</th>
<th>[more] [less]</th>
<th>Link to GO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modifier</td>
<td>Relation</td>
<td>GO ID</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence Ontology</th>
<th>[more] [less]</th>
<th>Link to SO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modifier Relation</td>
<td>SO ID</td>
<td>SO term</td>
</tr>
<tr>
<td>has_agent</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Disease</th>
<th>[more] [less]</th>
<th>Link to MIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relation</td>
<td>MIM ID</td>
<td>MIM name</td>
</tr>
<tr>
<td>agent_in</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Save means you are not finished.
Use the reference number to come back and finish.

Submit means you are done, your entry will be reviewed.

What to expect when you are done?

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.
**Annotation tool for PRO: RACE-PRO**

**Definition of the Protein Object**

1. Enter a UniProtKB identifier (*) or click here to insert a different sequence:

   MTREDDEL IFARNKSYG WPidayDEVDG CAMPAPTML PFFSGEHTH MPDPEDIFP 60
   YSNEKKEG YRREFEGING WEIINHNPH PFSQQAAKT QSNASSDEY BEETSYSER 120
   UTHRRGASN MIYTVNVDIT TPAARICBEQK UQFQQETTTA SYNLKSKQ 180
   GEAATPAK PFRGFRPEWM QSCPRDSDI ITREDPQNKQ QERQEQKRQ FKKDFEKQG 240
   ERDREDNFOFQEGVYER KNLRAKOTY STSE Spears DQKGFROUG QPQVTAHR 300
   NMLQRRHSG AQAAATQYQQ NLQ

   OR, click here to insert a different sequence:

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) [more]
   - Amino acid number: [ ]
   - Choose PTM: [ ]

4. Protein Object name (separate multiple names using ",") [more]
   - [ ]

5. Evidence Source (separate multiple IDs using ",") [more]
   - DB name: PubMed
   - IDs: 18706382

**Annotation of the Protein Object**

**Domain**

- Modifier Relation: Pfam ID, Pfam name, PMID
  - NOT has_part PF00855 PWWP domain 18706382

**Functional Annotation**

- Modifier Relation: GO ID, GO term, Interaction with, Relative to, PMID
  - located_in GO:000566 nucleus
  - participates_in GO:001541 negative regulation of transcription
    - 18706382

**Sequence Ontology**

- [add]

**Disease**

- [add]

**Comments:**

- cleavage site experiment in vitro.
FIGURE 5. The p38 cleavage fragment interferes with the transactivation potential of LEDGF/p75. A. Either 1 ng (1x) or 2 μg (2x) of pcDNA3-HA-LEDGF/p75 or pCruHa-p52 DNA were co-transfected with 0.2 μg (1x) or 0.4 μg (2x) of pCruHa-DN85 or empty vector (Vec), together with the pGL3-Hep2p75Luc reporter plasmid. HCT116 cells were lysed after 48 h and assayed for fold induction of luciferase activity. Columns, mean of three independent experiments done in quadruplicate; bars, SD. *, P < 0.05, one-way ANOVA with Bonferroni’s multiple comparison test (GraphPad Prism). B. Corresponding immunodot showing recombinant protein expression, detected with anti-HA antibody, 48 h after transfection in HCT116 cells. C. U2OS cells were transiently transfected with pCruRedLEDGF/p75, pCruRed-p52, or pCruHa-p38/DN85 and then grown in covergrips. After 48 h, recombinant protein expression in transfected cells was visualized by fluorescence microscopy. After fixation and permeabilization, pCruRedLEDGF/p75 and pCruRed-p52 were visualized directly, whereas HA-p38/DN85 was detected with primary rabbit anti-HA antibody with secondary Alexa 488-labeled goat antirabbit antibody. Nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI). D. Cells were co-transfected with pCruHa-p38/DN85 (0.5 μg) and pHcRedLEDGF/p75 (1 μg) or pHcRed-p52 (1 μg), stained with anti-HA antibodies, and visualized by fluorescence microscopy.
How to link to PRO

To link to PRO entry, please use the following URL:

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

where PRO:xxxxxxxxx is the corresponding PRO ID.

Example: http://pir.georgetown.edu/cgi-bin/pro/entry_pro?id=PRO:000000447

Contact us through

Pir help desk pirmail@georgetown.edu

Please specify that is a PRO question/comment

Or send me an email arighi@dbi.udel.edu

Please specify that is a PRO question/comment
PRO Consortium Team (so far…)

Principle Investigators
Cathy Wu (UD/GU-PIR)
Judith Blake (Jackson Lab-MGI)
Barry Smith (SUNY Buffalo-NCBO)

Co-Investigators, Curators and Developers
Darren Natale (GU-PIR)
Harold Drabkin (Jackson Lab-MGI)
Winona Barker (GU-PIR)
Uzoamaka Ugochukwu (GU-PIR)
Jian Zhang (GU-PIR)
Hongzhan Huang (UD-PIR)
Jules Nchoutmboube (UD-PIR)