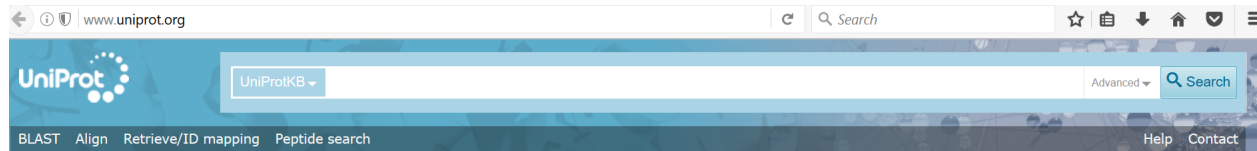# UniProt Guidelines

## URL
http://www.uniprot.org/

## Description
The UniProt Knowledgebase (UniProtKB) is a comprehensive resource for protein sequence and functional information with extensive cross-references to more than 120 external databases. Besides amino acid sequence and a description, it also provides taxonomic data and citation information.

## Guidelines
This database is to be used *if the name of the protein is known*. The following graphic shows the top of the UniProt homepage. You will type the name of your protein into the main search bar. Be sure that the "UniProtKB" option is selected in the dropdown menu to the left.



## Warning on Nomenclature
Some researchers use unofficial nomenclature to describe the proteins in their research. In this case, the name will probably not yield any viable results. An example of such an instance is PP2A (Protein Phosphatase 2).
If this is the case with your protein, you will need to find the proper name of the protein using either PubMed or Wikipedia. According to Wikipedia, the official identifier of this enzyme is "PPP2CA".

## Analyzing the Results
Continuing with the PPP2CA example, the search yields almost 200 results. To ensure that you have selected the correct one, you have to look at the "Protein Names", "Gene Names" and "Organisms" columns on the right side of the screen.

esults     ❓ About UniProtKB   🛒 Baske

| | Entry ⬍ | Entry name ⬍ | | Protein names ⬍   ⏩ | Gene names ⬍ | Organism ⬍ | Length ⬍ |
|---|---|---|---|---|---|---|---|
| ☐ | P04637 | P53_HUMAN | 📄 | Cellular tumor antigen p53 | TP53 P53 | Homo sapiens (Human) | 393 |
| ☐ | P67775 | PP2AA_HUMAN | 📄 | Serine/threonine-protein phosphatas... | PPP2CA | Homo sapiens (Human) | 309 |
| ☐ | P63331 | PP2AA_RAT | 📄 | Serine/threonine-protein phosphatas... | Ppp2ca | Rattus norvegicus (Rat) | 309 |
| ☐ | P63330 | PP2AA_MOUSE | 📄 | Serine/threonine-protein phosphatas... | Ppp2ca | Mus musculus (Mouse) | 309 |
| ☐ | P31749 | AKT1_HUMAN | 📄 | RAC-alpha serine/threonine... | AKT1 PKB RAC | Homo sapiens (Human) | 480 |

From the literature and the Wikipedia entry, we know that:

- the protein is a "...ubiquitous and conserved serine/threonine phosphatase..." (Wikipedia, 2016)
- The gene that makes this protein is the PPP2CA gene.
- The protein we want to annotate is found in humans.

The second entry is the one that we're most likely looking for, and we will use its unique identifier (P67775) to annotate in the BioModels repository.

**Using the Result to Help with the Rest of the Annotation**

If you click on the 'P67775' link, you are taken to this protein's record in UniProt. You will find that the record is split up into several sections. You will want to focus on the following sections (if they are available) to help you annotate reactions and compartments.

- Function
- Subcellular Location
- Cross-References

<div align="center">Function</div>

The function section contains a basic description of the protein along with: its function, Gene Ontology (GO) annotations, and Enzyme & Pathway database locations.

The GO annotations can be helpful for annotating biological processes. In this instance, there are over 20 biological processes associated with this protein. Most of them are tagged through the UniProtKB database.

## GO - Biological process[i]

- apoptotic process ✎ Source: UniProtKB ▾
- ceramide metabolic process ✎ Source: UniProtKB ▾
- inactivation of MAPK activity ✎ Source: UniProtKB ▾
- meiotic cell cycle ✎ Source: UniProtKB-KW
- mesoderm development ✎ Source: Ensembl
- mitotic nuclear envelope reassembly ✎ Source: Reactome
- negative regulation of cell growth ✎ Source: UniProtKB ▾
- negative regulation of epithelial to mesenchymal transition ✎ Source: BHF-UCL ▾
- negative regulation of tyrosine phosphorylation of Stat3 protein ✎ Source: UniProtKB ▾
- nuclear-transcribed mRNA catabolic process, nonsense-mediated decay ✎ Source: Reactome
- positive regulation of protein serine/threonine kinase activity ✎ Source: BHF-UCL ▾
- protein dephosphorylation ✎ Source: UniProtKB ▾
- regulation of cell adhesion ✎ Source: UniProtKB ▾
- regulation of cell differentiation ✎ Source: UniProtKB ▾
- regulation of DNA replication ✎ Source: UniProtKB ▾
- regulation of growth ✎ Source: UniProtKB ▾
- regulation of transcription, DNA-templated ✎ Source: UniProtKB ▾
- regulation of Wnt signaling pathway ✎ Source: UniProtKB ▾
- response to organic substance ✎ Source: UniProtKB ▾
- RNA splicing ✎ Source: UniProtKB ▾
- second-messenger-mediated signaling ✎ Source: UniProtKB ▾

Complete GO annotation...

If you scroll down further, you will see the Enzyme & Pathway subsection. This is where you will find the proper reaction annotations that link to its record in Reactome.

| Reactome[i] | R-HSA-113501. Inhibition of replication initiation of damaged DNA by RB1/E2F1. |
| --- | --- |
| | R-HSA-1295596. Spry regulation of FGF signaling. |
| | R-HSA-163685. Integration of energy metabolism. |
| | R-HSA-163767. PP2A-mediated dephosphorylation of key metabolic factors. |
| | R-HSA-180024. DARPP-32 events. |
| | R-HSA-195253. Degradation of beta-catenin by the destruction complex. |
| | R-HSA-196299. Beta-catenin phosphorylation cascade. |
| | R-HSA-198753. ERK/MAPK targets. |
| | R-HSA-202670. ERKs are inactivated. |
| | R-HSA-2465910. MASTL Facilitates Mitotic Progression. |
| | R-HSA-2467813. Separation of Sister Chromatids. |
| | R-HSA-2500257. Resolution of Sister Chromatid Cohesion. |
| | R-HSA-2995383. Initiation of Nuclear Envelope Reformation. |
| | R-HSA-389513. CTLA4 inhibitory signaling. |
| | R-HSA-432142. Platelet sensitization by LDL. |
| | R-HSA-4641262. Disassembly of the destruction complex and recruitment of AXIN to the membrane. |
| | R-HSA-5339716. Misspliced GSK3beta mutants stabilize beta-catenin. |
| | R-HSA-5358747. S33 mutants of beta-catenin aren't phosphorylated. |
| | R-HSA-5358749. S37 mutants of beta-catenin aren't phosphorylated. |
| | R-HSA-5358751. S45 mutants of beta-catenin aren't phosphorylated. |
| | R-HSA-5358752. T41 mutants of beta-catenin aren't phosphorylated. |
| | R-HSA-5467337. APC truncation mutants have impaired AXIN binding. |
| | R-HSA-5467340. AXIN missense mutants destabilize the destruction complex. |
| | R-HSA-5467348. Truncations of AMER1 destabilize the destruction complex. |
| | R-HSA-5663220. RHO GTPases Activate Formins. |
| | R-HSA-5673000. RAF activation. |
| | R-HSA-5675221. Negative regulation of MAPK pathway. |
| | R-HSA-6804757. Regulation of TP53 Degradation. |
| | R-HSA-6811558. PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling. |
| | R-HSA-68877. Mitotic Prometaphase. |
| | R-HSA-69231. Cyclin D associated events in G1. |
| | R-HSA-69273. Cyclin A/B1 associated events during G2/M transition. |
| | R-HSA-70171. Glycolysis. |
| | R-HSA-975957. Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC). |

Choose the reaction that best fits the one described in the paper. For example, if you were to use one of these annotations to put in the BioModels database, it would look something like this: isVersionOf R-HSA-163767: "PP2A-Amediated dephosphorylation of key metabolic factors"

## Subcellular Location

This section contains all of the cellular components that are associated with the protein. The tags can come from a variety of databases, so you will need to choose which ones are interoperable with the BioModels repository.

## Cross-References

The cross-references section holds the additional databases that this protein is found under. For the PP2A protein, one area of interest is the "Genome Annotation Databases" section which contains the KEGG tag (hsa:5515). This tag links to the protein's record in KEGG. It's a good idea to check this section when searching other proteins, because there may be tags that could be used in the BioModels repository.