

# TEACHING THE GENOME GENERATION

*UniProt Tutorials*



## UniProt Tutorials

Use the following menu to link to a specific topic within this tutorial:

- A. [UniProt Background and Home Page](#)
- B. [Searching for a Protein](#)
- C. [Navigating a UniProtKB Entry](#)
  - i. [Protein Function](#)
  - ii. [Protein Features](#)

### Reference:

The UniProt Consortium. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1):D480–D489. doi: 10.1093/nar/gkaa1100

## A. UniProt Background and Home Page

The **Universal Protein Resource (UniProt)** contains information on protein sequences and functions. There are several different databases within UniProt. This tutorial will focus on the UniProt Knowledgebase (UniProtKB), which contains functional information about proteins from a wide variety of species, including humans. See also the [UniProt Tutorial Series: Introduction](#) video.

1. Navigate to [UniProt](#). The homepage contains menu tabs that direct to different resource functions (1), a search bar for searching the databases (2), and links to each of the UniProt databases (3).

The screenshot shows the UniProt homepage. At the top, a navigation menu (1) includes UniProt, BLAST, Align, Peptide search, ID mapping, and SPARQL. The main search area (2) features the heading "Find your protein", a search bar with "UniProtKB" selected, and a "Search" button. Below the search bar are examples: "Insulin, APP, Human, P05067, organism\_id:9606". At the bottom, four database tiles (3) are displayed: "Proteins UniProt Knowledgebase" (with sub-tiles for Reviewed (Swiss-Prot) 568,363 and Unreviewed (TrEMBL) 229,928,140), "Species Proteomes" (Protein sets for species with sequenced genomes from across the tree of life), "Protein Clusters UniRef" (Clusters of protein sequences at 100%, 90% & 50% identity), and "Sequence Archive UniParc" (Non-redundant archive of publicly available protein sequences seen across different databases).

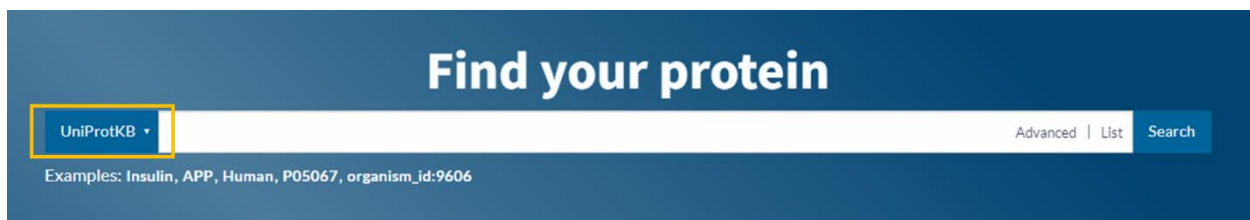
**Figure 1. UniProt Homepage.** The UniProt homepage includes (1) menu tabs, (2) a search bar, and (3) links to databases.

2. In this tutorial, we will search UniProtKB for the entry on the human protein Angiotensin-converting enzyme (ACE).

## B. Searching for a Protein

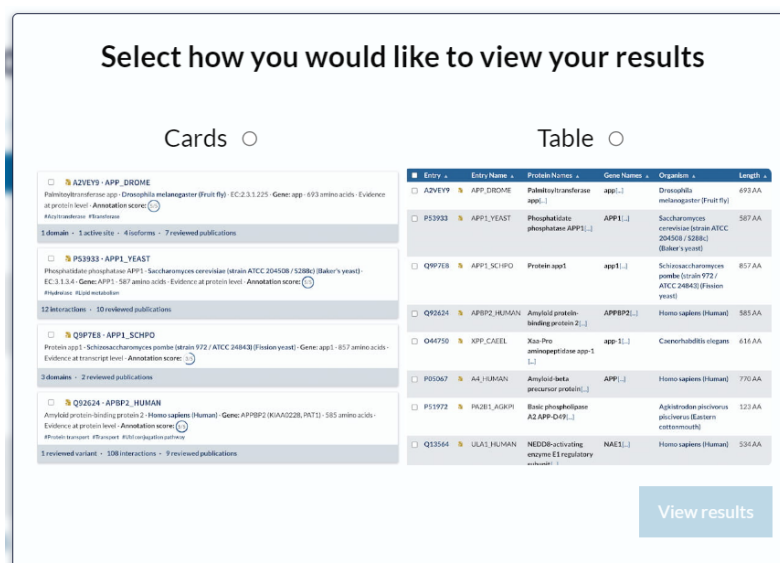
This section will demonstrate how to search for a protein of interest, using human Angiotensin-converting enzyme (ACE) as an example. This tutorial will begin on the home page of [UniProt](#). See also the [UniProt Tutorial Series: Searching for a Protein](#) video.

1. To search for a protein, locate the search bar labeled **Find your protein** on the home page of UniProt.
2. You can select which database you would like to search using the dropdown menu on the left of the search bar. For this tutorial, we will search UniProtKB, which is the default option.



**Figure 2. Search Bar.** The search bar on the UniProt homepage allows you to select which database to search.

3. To search for the human ACE protein, type Human ACE into the search bar and click **Search**.
4. When the results load, you may see an option to choose between cards or a table to display the results. Select one by clicking the circle next to the option. If you have used UniProt before, the browser may automatically use a previously selected view option. If you select **table** view, proceed to step 5. If you select **card** view, proceed to step 7.



**Figure 3. Result View Selection.** Select either cards or table view for displaying search results in UniProt.

- In the table view, check the **Entry Name** column. UniProt entry names typically consist of the name of the protein and the name of the organism, styled as PROTEIN\_ORGANISM. For this tutorial, locate ACE\_HUMAN. You can confirm you have found the correct entry by checking the **Protein Names** and **Organism** columns.

## UniProtKB 173 results

BLAST Align Map IDs Download Add View: Cards Table Customize columns Share

Entry	Entry Name	Protein Names	Gene Names	Organism	Length
<input type="checkbox"/> P12821	ACE_HUMAN	Angiotensin-converting enzyme[...]	ACE, DCP, DCP1	Homo sapiens (Human)	1,306 AA
<input type="checkbox"/> Q9BYF1	ACE2_HUMAN	Angiotensin-converting enzyme 2 [...]	ACE2, UNQ868/PRO1885	Homo sapiens (Human)	805 AA
<input type="checkbox"/> Q9GLN7	ACE_PANTR	Angiotensin-converting enzyme[...]	ACE, DCP1	Pan troglodytes (Chimpanzee)	1,304 AA
<input type="checkbox"/> Q5EGZ1	ACE2_RAT	Angiotensin-converting enzyme 2 [...]	Ace2	Rattus norvegicus (Rat)	805 AA

**Figure 4. Results Table.** In the table view, each row contains information about a different protein entry.

- Click on the link in the **Entry** column to access the protein entry. For this tutorial the human ACE Entry is labeled **P12821**.
- In the cards view, check the card title. UniProt entry names typically consist of the name of the protein and the name of the organism, styled as PROTEIN\_ORGANISM. For this tutorial, locate ACE\_HUMAN. You can confirm you have found the correct entry by checking the **protein name** (Angiotensin-converting enzyme) and **organism** (Homo sapiens) in the description below the title.

## UniProtKB 178 results

BLAST Align Map IDs Download Add View: Cards Table Share

<input type="checkbox"/> <b>P12821 · ACE_HUMAN</b>
Angiotensin-converting enzyme · Gene: ACE (DCP, DCP1) · Homo sapiens (Human) · EC:3.4.15.1 · 1306 amino acids · Evidence at protein level · Annotation score: (5/5)
#Calmodulin-binding #Carboxypeptidase #Hydrolase #Metalloprotease #Protease
1 PTM · 24 reviewed variants · 4 active sites · 4 isoforms · 1 interaction · 4 diseases · 63 3D structures · 73 reviewed publications
<input type="checkbox"/> Q9BYF1 · ACE2_HUMAN
Angiotensin-converting enzyme 2 · Gene: ACE2 · Homo sapiens (Human) · EC:3.4.17.23 · 805 amino acids · Evidence at protein level · Annotation score: (5/5)
#Carboxypeptidase #Host cell receptor for virus entry #Hydrolase #Metalloprotease #Protease #Receptor #Host-virus interaction
2 PTMs · 4 reviewed variants · 2 active sites · 2 isoforms · 31 interactions · 177 3D structures · 69 reviewed publications
<input type="checkbox"/> Q9GLN7 · ACE_PANTR
Angiotensin-converting enzyme · Gene: ACE (DCP1) · Pan troglodytes (Chimpanzee) · EC:3.4.15.1 · 1304 amino acids · Inferred from homology · Annotation score: (5/5)
#Calmodulin-binding #Carboxypeptidase #Hydrolase #Metalloprotease #Protease
1 PTM · 4 active sites · 2 isoforms · 1 reviewed publication

**Figure 5. Results Cards.** In the card view, each card is titled with the protein entry ID and entry name.

- Click on the title (**P12821 · ACE\_HUMAN**) to access the protein entry.

## C. Navigating a UniProtKB Entry

This section will focus on how to find protein structure and function information within a UniProtKB entry. See also the UniProt Tutorial Series videos [Navigating a UniProtKB Protein Entry](#) and [Finding Information on Protein Function](#).

1. Each protein entry contains information about protein structure, function, and expression. At the top of each entry is a summary header (1), which provides basic information about the protein, including its name, length, and associated gene.

The screenshot shows the UniProtKB entry for P12821 (ACE\_HUMAN). The entry is divided into three main sections:

- Section 1 (Summary Header):** Located at the top right, it contains key information:
 

Protein <sup>i</sup>	Angiotensin-converting enzyme	Amino acids	1306
Status <sup>i</sup>	UniProtKB reviewed (Swiss-Prot)	Protein existence <sup>i</sup>	Evidence at protein level
Organism <sup>i</sup>	Homo sapiens (Human)	Annotation score <sup>i</sup>	5/5
Gene <sup>i</sup>	ACE (DCP, DCP1)		
- Section 2 (Navigation Menu):** Located below the summary header, it contains tabs for: Entry (selected), Feature viewer, Publications, External links, and History.
- Section 3 (Table of Contents):** Located on the left side, it lists various sections of the entry: Function, Names & Taxonomy, Subcellular Location, Disease & Variants, PTM/Processing, Expression, Interaction, Structure, Family & Domains, Sequence & Isoforms, and Similar Proteins.

Below the navigation menu, there are links for BLAST, Align, Download, Add, Community curation (1), Add a publication, and Entry feedback. The 'Function' section is expanded, showing the following text:

**Function<sup>i</sup>**  
 Dipeptidyl carboxypeptidase that removes dipeptides from the C-terminus of a variety of circulating hormones, such as angiotensin I, bradykinin or enkephalins, thereby playing a key role in the regulation of blood pressure, electrolyte homeostasis or synaptic plasticity (PubMed:2558109, PubMed:4322742, PubMed:7683654, PubMed:7523412, PubMed:15615692, PubMed:20826823).

**Figure 6. UniProt Entry.** A UniProt entry contains (1) a summary, (2) a navigation menu, and (3) a table of contents.

2. Beneath the header is a navigation menu (2). Tabs on this menu direct to additional features of UniProt that provide information about the selected protein. This tutorial will focus on the **Entry** tab, which is selected by default.
3. A clickable table of contents (3) on the left enables you to quickly navigate to different sections of the entry. You can also navigate through the entry by scrolling in your browser. The sections of a UniProt entry include:

### Function

Provides a summary of the known biological function of the protein, including information about structural features.

### Names & Taxonomy

Provides protein and gene names, plus information about the organism from which the protein comes.

### Subcellular Location

Provides information about where in the cell the protein can typically be found.

**Disease & Variants**

Provides information about any known diseases or phenotypes associated with the protein, along with details on protein variants.

**PTM/Processing**

Provides details on post-translational modification (PTM) and processing of the protein.

**Expression**

Provides data on the mRNA and protein expression levels.

**Interaction**

Provides information on protein quaternary structure and on interactions with other proteins.

**Structure**

Provides information on secondary and tertiary structure, if known.

**Family & Domains**

Provides information on similarity to other proteins.

**Sequence & Isoforms**

Provides the protein sequence and any known isoforms, or alternate versions, of the protein sequence.

**Similar proteins**

Provides links to proteins that are similar in sequence.

*i. Protein Function*

1. The first section of the main entry page, titled **Function**, includes a description of the protein function.
2. Throughout the description, UniProt links to references of scientific journal articles and reviews that provide evidence for the functional information. If you would like to learn more, you can click the linked text to be redirected to the relevant article.
3. At the end of the text, there may be tags such as “# Publications” or “By Similarity”. These tags indicate that the functional information was derived from a number of scientific papers or based on sequence similarity to a protein in another organism. Clicking a tag provides more information about, and links to, the relevant sources.

## ii. Protein Features

UniProt entries contain information on several different types of protein features, including functional features, post-translational modification sites, and protein variants. Feature information can be accessed either within the relevant sections of the **Entry** or through the **Feature viewer**. This tutorial will focus on locating the functional features within the main UniProt **Entry**.

1. Scroll down and locate the **Function** section of the **Entry**.
2. Find the **Features** heading within the **Function** section.

**Features**  
Showing features for region<sup>1</sup>; binding site<sup>2</sup>; active site<sup>3</sup>; site<sup>4</sup>.

1 1 100 200 300 400 500 600 700 800 900 1,000 1,100 1,200 1,300 1306

2

TYPE	ID	POSITION(S)	DESCRIPTION
-- Select --	5		
▶ Region	30-630		Peptidase M2 1 <span>BLAST</span> <span>Add</span>
▶ Binding site	231		chloride 1 (UniProtKB   ChEBI <a href="#">↗</a> ) <span>1 Publication</span>
▶ Binding site	390		Zn <sup>2+</sup> 1 (UniProtKB   ChEBI <a href="#">↗</a> ); catalytic <span>3 Publications</span> <span>Combined Sources</span>
▶ Active site	391		Proton acceptor 1 <span>1 Publication</span> <span>2 Publications</span>
▶ Binding site	394		Zn <sup>2+</sup> 1 (UniProtKB   ChEBI <a href="#">↗</a> ); catalytic <span>3 Publications</span> <span>Combined Sources</span>
▶ Binding site	414		Zn <sup>2+</sup> 1 (UniProtKB   ChEBI <a href="#">↗</a> ); catalytic; In isoform P12821-3; <span>1 Publication</span> <span>Combined Sources</span>
▶ Binding site	418		Zn <sup>2+</sup> 1 (UniProtKB   ChEBI <a href="#">↗</a> ); catalytic <span>3 Publications</span> <span>Combined Sources</span>
			Zn <sup>2+</sup> 1 (UniProtKB   ChEBI <a href="#">↗</a> ); catalytic; In isoform P12821-3; <span>1 Publication</span>

4 Expand table

**Figure 7. Protein Features.** The Features table of a UniProt entry details important protein features, such as the active site.

3. Notice the length of the protein is shown in grey (1). For Human ACE, the diagram shows amino acids 1 through 1306.
4. Underneath the grey bar, colored shapes (2) identify sections of the protein that are important for its function, also known as functional features. These features are also listed in the table below the grey bar (3). To view the whole table without having to scroll, click the **Expand table** button (4).
5. ACE is an enzyme, so its features include enzymatic active sites. Other features include binding sites, cleavage sites, and other regions of interest. To filter for a particular type of feature, use the **TYPE** dropdown menu (5).