

Basic Protein Query: Searching for Unique Identifiers

Unique identifiers can be accession numbers, which apply to a complete sequence record, or sequence identification numbers, which apply to the individual sequences within a record.

The format of accession numbers varies, depending upon the source database. Some descriptions and examples of typical unique identifiers are given below (adapted from NCBI).

Type of Record

Gene ID

Sample Accession Format

This refers to the NCBI Entrez Gene ID. The identifier that is assigned is an integer and is species specific. For genomes that had been represented in NCBI LocusLink, the Gene ID is the same as the Locus ID. The Gene ID is reported in RefSeq records as a ?db_xref? (e.g., /db_xref=?Gene ID:856646,? in GenBank format).

Type of Record

UniProt Accession Number

Sample Accession Format

There are a total of six characters.

Use capital letters. Character/Format

1[O,P,Q]

2[0-9]

3[A-Z,0-9]

4[A-Z,0-9]

5[A-Z,0-9]

6[0-9]e.g.: P12345 and Q9JJS7

After February 7, 2006:

UniProt wants to make sure that the space of available accession numbers is adequate to deal with the large expansion in the number of protein sequences and is therefore going to extend the current scheme by allowing the first character of the accession number to be any of the 26 letters (instead of only O, P and Q).



ImmPort Protein Query

Basic Protein Query: Searching for Unique Identifiers

Type of Record UniProt Accession Number	Sample Accession Format To avoid assigning accession numbers identical to those that have been used by the international nucleotide sequence database, the extension of the first position goes along with a restriction in the third position, which can only be a letter. The new format for UniProtKB accession numbers will therefore be: Character/Format 1[A-N,R-Z] 2[0-9] 3[A-Z] 4[A-Z,0-9] 5[A-Z,0-9] 6[0-9] e.g.: B1J345 and R9JJS7
Type of Record RefSeq Protein Accession	Sample Accession Format Two capital letters (NP), an underscore bar, and six digits, e.g.: NP_000483.



Reference Data / Protein Query

[Home](#) | [Genes](#) | [Proteins](#) | [MHC Alleles](#) | [Pathways](#) | [Protein Networks](#) | [SNPs](#) | [ImmPort Gene Lists](#) ▾ | [Immunologically Related Genes](#) | [Download](#) ▾ | [Reference Data History](#)

Fields marked with an asterisk * are required.

Search Options

Search Type*	<input type="text" value="Protein Name"/>
Search Option*	<input type="text" value="Like"/>
Search Text* (Comma delimited, 256 max chars)	<input type="text" value="syk"/>
Search Species* (Multiple)	<input type="text" value="Homo sapiens"/> Mus musculus Rattus rattus Rattus norvegicus Gallus gallus Drosophila melanogaster Macaca mulatta Select All

Enter the protein query parameters and click submit to view the details

Results Per Page:

Results per page has a max of 100 items

Protein Query

Fields marked with an asterisk * are required.

Details of the filters on which to do a protein query

Search Type*

Gene Symbol
Gene Symbol
Gene Name
Gene ID
Chromosome Region
NCBI mRNA Accession
NCBI Protein Accession

Search type used in the query

Search Option*

Like
Like
Exact Match
Start With
Like

Search option operator used in the query

Search Text*
(Comma delimited, 256 max chars)

il20

Search text used in the query like a gene of interest

Search Species*
(Multiple)

Homo sapiens
Mus musculus
Rattus rattus
Rattus norvegicus
Gallus gallus
Drosophila melanogaster
Macaca mulatta
[Select All](#)

Species studied can be selected individually or in multiples using the Ctrl Key. OR click 'Select All'

Search Human Build*
(for Homosapiens)

Build 36.1, hg18
Build 36.1, hg18

Human build

Results Per Page

25

Results per page up to a maximum of 100 and click 'Submit'