

PTM SEARCH TUTORIAL

PTM search allows users to retrieve PTMs in a specific amino acid sequence context and/or combinations of PTMs. This is possible thanks to a regular expression search that allows to search for ambiguity in PTM type and amino acids. As example we consider the search 'K(Ub)X{1,4}S(Ph)', *i.e.* a ubiquitinated lysine, 'K(Ub)', followed maximal five amino acids downstream, X{1,4}, by a phosphorylated serine 'S(Ph)'. For more details regarding the regular expression please consult the extensive help section below on the PTM Search page. Besides inputting the PTM Search query, several advanced options can be fine-tuned. For instance, resulting PTMs can be required to be resorting from peptides matching proteins encoded from a single gene, restrict the search to a single plant species and the eventually displayed sequence window length. In addition, there are advanced options that require PTMs to reside in specific protein regions. For instance, PTMs can be required to reside in a specific InterPro protein domain, by inputting the InterPro identifier, or match a UniProtKB site-specific annotation. For the site annotations, a selection to four types: active site, binding site, metal ion-binding site and site. For more info regarding these site annotations we refer to the UniProtKB help section (https://www.uniprot.org/help/sequence_annotation).

The screenshot shows the PTM Search interface with several annotations in pink boxes:

- Check: only representative protein, no isoforms**: Points to the **Representative Only** checkbox.
- Check: peptides match only proteins from single gene**: Points to the **Peptides Unique** checkbox.
- Restrict to species**: Points to the **Species** dropdown menu.
- Sequence window length for results table**: Points to the **Sequence Length** input field.
- Enter PTM sequence to search, e.g. 'K(Ub)X{1,4}(Ph)S'**
See below detailed info wildcards and operators!: Points to the **Search Expression** input field.
- PTM inside InterPro protein domain**: Points to the **Domain** input field.
- PTMs match UniProtKB site annotation**
E.g. active site, metal-binding, ...: Points to the **Sites** selection options.
- SEARCH**: Points to the **SEARCH** button.
- Search PTMs fulfilling specified criteria**: A summary box at the bottom.

The interface includes a navigation bar with links: HOME, PROTEIN SEARCH, PTM SEARCH (highlighted), PTM BLAST, EXPERIMENTS, PTM INFO, BROWSE/DOWNLOAD, SUBMIT, and RESOURCES. The main form contains a **Search Expression** field, a **HIDE ADVANCED OPTIONS** button, and several filter sections: **Representative Only** (checked), **Peptides Unique** (unchecked), **Species** (set to **<ANY>**), **Sequence Length** (set to 10), **Domain** (empty), **IPR** (empty), and **Sites** (with options: Active Site, Binding Site, Metal Ion-binding Site, Site). A warning message at the bottom states: "Be aware that selecting any of the domain/site options may cause the search to run considerably more slowly, so make sure that your search terms are as narrow as possible before submitting. If your search returns more than 500 results without domain/site options set, do not run again with these options".

After clicking the 'SEARCH' button, PTM Search will be executed and retrieve matching PTM sequences. Note that for highly ambiguous searches, e.g. many '(xx)' as PTMs or 'X' as amino acid, the PTM search might take some time. This is especially true for PTM searches restricted to protein domains or UniProtKB annotated sites. We are currently optimizing the speed for these last search options. After the PTM Search is finished, a search result table will appear on the bottom of the page. This table is organized across multiple pages and the results per page can be customized. The results are organized per protein and above the total amount of proteins that contain the searched PTM sequence are displayed. Some proteins might have multiple PTM sequences fulfilling the rules. For instance, for our search, AT1G06070.1 a phosphoserine is found 2 and 5 positions after a ubiquitinated lysine. The respective PTM positions within the protein are displayed for the first and last PTM specified in the PTM search. In case a single PTM was searched, logically only one PTM position is returned. In the returned PTM sequence window, which length can be customized (see above), all PTMs are highlighted according PTM type. In case if the PTM sequence matches proteins encoded from multiple gene loci, these will be flagged. Protein identifiers in the left column provide a link to the respective protein overview page.

Number of proteins matching the PTM Search: **171 matching proteins found**
 'K(Ub)X{1,4}S(Ph)', i.e. ubiquitinated Lys followed by phosphorylated Ser 1 to 5 positions downstream

SHOW HELP

Search Results

Page of 7 Results per page:

In case > 2 PTMs, the positions are shown for first and last PTM specified in PTM search
 Here: K(Ub) Here: S(Ph)

Protein ID	Description	Species	First PTM pos	Last PTM pos	Surrounding Sequence
AT1G02780.1	Ribosomal protein L19e family protein	ath	153	156	HKSKAEKAREKTLSDQFEAKRAKNI
AT1G04410.1	Lactate/malate dehydrogenase family protein	ath	237	242	TVQQRGAATIKARKIKSAIQAASSAK
AT1G05190.1	Ribosomal protein L6 family	ath	87	90	LTYPREVELTKEEESGFLRVKKTVE
AT1G06070.1	Basic-leucine zipper (bZIP) transcription factor family protein	ath	408	410	LRRRPMKPGQKESVTSKDREIPL
			408	413	LRRRPMKPGQKESVTSKDREIPLTKD
AT1G07790.1	Histone superfamily protein	ath	110	115	ESSKLARYNKKPTIISREIQTAVRLV
AT1G08370.1	decapping 1	ath	143	147	ILSAYSVWQPKASSSKSEFEELE

Link to protein overview

PTM sequence matches proteins encoded by multiple gene loci

PTMs are highlighted according to PTM type(s)
 All protein PTMs are shown for those positions, not only the searched PTMs