**ProteoMapper Web tutorial**

Luis Mendoza
Institute for Systems Biology
Last Revision: May 25, 2021

**About**

This tutorial covers basic usage of the **ProteoMapper** tools through the online resource at *PeptideAtlas*. Please contact the author if you have any questions or have any issues with this tutorial.

**Introduction**

**ProteoMapper** is a set of tools that enable the mapping of peptide sequences to protein sequences using a full reverse index of segments. The software supports protein sequence variants (specified via the *PEFF* keyword *VariantSimple*), as well as wildcard and fuzzy mapping, multiple output formats, and is very quick.

For more information, visit: http://www.tppms.org/tools/pm/

The online resource at *PeptideAtlas* allows a user to investigate the mapping of one of more peptide sequences (up to 5,000) against the proteomes of various organisms; for human, one is able to map against the *neXtProt* database, or the full *THISP* one used by *PeptideAtlas*.

**Prerequisites and URL**

For this tutorial, a web browser and internet connection are all that is needed. Please point your browser to http://www.peptideatlas.org/map/ .

**Single Peptide Mapping**

1. Find the protein mapping of peptide **PPLPKSR** against **the Human neXtProt + contaminants** database. Select this database and enter the peptide sequence into the form; leave all other settings unchanged. Press “Search!” and wait a few seconds for results to appear.

Examine the results.

* How many proteins does it map to? Note that some of these protein entries are isoforms of each other.
* Does it map “cleanly” – that is, without SAAVs – to any proteins? You can find the “nSubs” and corresponding graphical bar useful.
* How about using one SAAV? Two SAAVs?
1. For the same sequence, examine its mappings against the **All Human PeptideAtlas** database.

Notice that there are more hits than before; that is because the *THISP* database used by *PeptideAtlas* contains extra sequences in addition to the *neXtProt* entries. You can find more information by clicking on the Column Descriptions link at the top of the results.

Also, note that the top portion of the results is reserved for *neXtProt* mappings; convince yourself that they match the results from step 1 above.

For the rest of the results, find instances of identical sequences, as well as those that appear very similar to each other. Follow the links to *PeptideAtlas* for further study.

1. Is this same sequence found in any of the other organism proteomes?

**Fuzzy Sequence Mapping**

1. Use the fuzzy mapping function to map a peptide sequence from a de novo search. Let’s say that you ran a de novo algorithm on a spectrum, and got the following sequence as the best match: **EDDSLSPASANDDK**
* Does this sequence map to any proteins in **All Human PeptideAtlas**?
* Use the fuzzy mapping setting of one unknown amino acid and a mass tolerance of 0.01. Are there any matches?
* Run again with two unknown amino acids. Are there any matching proteins? What was the actual peptide sequence that was matched to the best matched *neXtProt* protein? Do you expect this to be common when doing de novo sequencing? Why/why not?
* By default, the web tool does not display redundant matches in order to simplify the output. You can expand or re-hide these extra results by clicking on the Show All Redundant Mappings toggle link.

**Mapping Multiple Sequences**

1. By clicking on the single peptide or peptide list link at the top of the form, you can toggle between a single sequence and a list input. Note, however, that fuzzy mapping is only available in single sequence mode.

Find a suitable list of peptides to paste into the Peptide Sequences input box, and run it against the most appropriate database. Notice that those sequences that cannot be mapped are highlighted in the results.