

Part I:

1. What are the identities of samples 1 and 2, based on the results of at least two of the search engines mentioned below?

Test Sample 1: Exportin-1 in YEAST / CAH2_BOVIN, Carbonic anhydrase 2 OS=Bos taurus & Protein FAM229A OS=Homo sapiens
Test Sample 2: Glycogen phosphorylase, muscle form - in RABIT / PYGM_RABIT, Glycogen phosphorylase, muscle form

2. Are there any differences in the results that the search engines gave back and if yes, discuss why you might be getting different results.

Test sample one had two different results from the search engines. This could be because of sample length. Sample 2 had more mass data than sample 1, meaning that it's possible to make a more concise match to the correct protein in the database.

3. Why were the samples carboxymethylated?

The samples are carboxymethylated to modify the hydroxyl groups in the proteins. This allows enzymes to digest the protein easier and promotes separation

4. Which peaks would shift if the sample had been carbamidomethylated and why? Yes, because I think the proteins would have stuck together, making it harder to distinguish unique proteins

Part II: Exploring bio.tools and selecting a new tool

On this website, we can find various open-source tools for proteomics. Tools can be web applications, desktop applications, programming libraries, frameworks, etc. We are going to walk through how to use these tools. By taking an example of the psm_utils library, we will use its functionalities.

Installing in Python

Installing it in python is much more interesting to use. There are a lot of functionalities that we can harness from this tool through code.

In the code cell below, install the psm_utils library using the pip command provided in the documentation. (Hint: Use ! sign before the pip command - because we can only install packages in the command line, and ! sign says that the particular cell should be run in the command line)

```
#Write your code in this cell
!pip install psm-utils
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting psm-utils
  Downloading psm_utils-0.2.1-py3-none-any.whl (49 kB)
    49.0/49.0 KB 2.9 MB/s eta 0:00:00
Collecting pyteomics>=4
  Downloading pyteomics-4.5.6-py2.py3-none-any.whl (232 kB)
    232.1/232.1 KB 24.0 MB/s eta 0:00:00
Requirement already satisfied: pandas in /usr/local/lib/python3.8/dist-packages (from psm-utils) (1.3.5)
Requirement already satisfied: pydantic in /usr/local/lib/python3.8/dist-packages (from psm-utils) (1.10.5)
Requirement already satisfied: click in /usr/local/lib/python3.8/dist-packages (from psm-utils) (8.1.3)
Requirement already satisfied: numpy in /usr/local/lib/python3.8/dist-packages (from psm-utils) (1.22.4)
Requirement already satisfied: lxml in /usr/local/lib/python3.8/dist-packages (from psm-utils) (4.9.2)
Collecting rich
  Downloading rich-13.3.1-py3-none-any.whl (239 kB)
    239.0/239.0 KB 26.2 MB/s eta 0:00:00
Collecting psims
  Downloading psims-1.2.3-py3-none-any.whl (14.7 MB)
    14.7/14.7 MB 56.8 MB/s eta 0:00:00
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.8/dist-packages (from pandas->psm-utils) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.8/dist-packages (from pandas->psm-utils) (2022.7.1)
Requirement already satisfied: sqlalchemy in /usr/local/lib/python3.8/dist-packages (from psims->psm-utils) (1.4.46)
Requirement already satisfied: six in /usr/local/lib/python3.8/dist-packages (from psims->psm-utils) (1.15.0)
Requirement already satisfied: typing-extensions>=4.2.0 in /usr/local/lib/python3.8/dist-packages (from pydantic->psm-utils) (4.5.0)
Collecting markdown-it-py<3.0.0,>=2.1.0
  Downloading markdown_it_py-2.2.0-py3-none-any.whl (84 kB)
    84.5/84.5 KB 10.0 MB/s eta 0:00:00
Collecting pygments<3.0.0,>=2.14.0
  Downloading Pygments-2.14.0-py3-none-any.whl (1.1 MB)
    1.1/1.1 MB 64.2 MB/s eta 0:00:00
Collecting mdurl~=0.1
  Downloading mdurl-0.1.2-py3-none-any.whl (10.0 kB)
Requirement already satisfied: greenlet!=0.4.17 in /usr/local/lib/python3.8/dist-packages (from sqlalchemy->psims->psm-utils) (2.0.2)
Installing collected packages: pyteomics, pygments, mdurl, psims, markdown-it-py, rich, psm-utils
  Attempting uninstall: pygments
    Found existing installation: Pygments 2.6.1
    Uninstalling Pygments-2.6.1:
      Successfully uninstalled Pygments-2.6.1
```

ERROR: pip's dependency resolver does not currently take account all the packages that are installed. This behaviour is the source of the following errors. To avoid this problem, please see https://pip.pypa.io/en/latest/topics/dependency-resolution/#diverging-requirements-versions.
ERROR: pip's dependency resolver does not currently take account all the packages that are installed. This behaviour is the source of the following errors. To avoid this problem, please see https://pip.pypa.io/en/latest/topics/dependency-resolution/#diverging-requirements-versions.
ipython 7.9.0 requires jedi>=0.10, which is not installed.
Successfully installed markdown-it-py-2.2.0 mdurl-0.1.2 psims-1.2.3 psm-utils-0.2.1 pygments-2.14.0 pyteomics-4.5.6 rich-13.3.1

Importing the Libraries

We have successfully installed the psm_utils package. To use it, we have to import it. Write the code to import the required library.

Note: We need to import the psm_utils library and the io object. See the image (from the documentation) below if you need help understanding. We need the io object because it contains APIs for reading, writing and converting PSM files.

Hint: We can use the import command to import the psm_utils package. But remember, we also have to import the io object from the psm_utils package. Use the "from _ import _" command to import the io from psm_utils.

PYTHON API REFERENCE

psm_utils

psm_utils.io

```
#Write your code in this cell
from psm_utils import Peptidoform, PSM, PSMList
from psm_utils.io import read_file
```

```
/usr/local/lib/python3.8/dist-packages/psims/mzmlb/writer.py:17: UserWarning: hdf5plugin is missing! Only the slower GZIP compression scheme will be available.
warnings.warn()
```

```
#ignore this cell but run it
!pip install ipywidgets
import ipywidgets
```

```
Requirement already satisfied: traitlets>=4.3.1 in /usr/local/lib/python3.8/dist-packages (from ipywidgets) (5.7.1)
Requirement already satisfied: ipython>=4.0.0 in /usr/local/lib/python3.8/dist-packages (from ipywidgets) (7.9.0)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.8/dist-packages (from ipywidgets) (3.0.5)
Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.8/dist-packages (from ipywidgets) (5.3.4)
Requirement already satisfied: ipython-genutils<=0.2.0 in /usr/local/lib/python3.8/dist-packages (from ipywidgets) (0.2.0)
Requirement already satisfied: widgetsnbextension<=3.6.0 in /usr/local/lib/python3.8/dist-packages (from ipywidgets) (3.6.2)
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.8/dist-packages (from ipykernel>=4.5.1->ipywidgets) (6.1.12)
Requirement already satisfied: tornado>=4.2 in /usr/local/lib/python3.8/dist-packages (from ipykernel>=4.5.1->ipywidgets) (6.2)
Requirement already satisfied: decorator in /usr/local/lib/python3.8/dist-packages (from ipython>=4.0.0->ipywidgets) (4.4.2)
Requirement already satisfied: pexpect in /usr/local/lib/python3.8/dist-packages (from ipython>=4.0.0->ipywidgets) (4.8.0)
Requirement already satisfied: backcall in /usr/local/lib/python3.8/dist-packages (from ipython>=4.0.0->ipywidgets) (0.2.0)
Requirement already satisfied: prompt-toolkit<2.1.0,>=2.0.0 in /usr/local/lib/python3.8/dist-packages (from ipython>=4.0.0->ipywidgets) (2.0.10)
Collecting jedi>=0.10
  Downloading jedi-0.18.2-py2.py3-none-any.whl (1.6 MB)
    1.6/1.6 MB 50.3 MB/s eta 0:00:00
Requirement already satisfied: setuptools>=18.5 in /usr/local/lib/python3.8/dist-packages (from ipython>=4.0.0->ipywidgets) (57.4.0)
Requirement already satisfied: pygments in /usr/local/lib/python3.8/dist-packages (from ipython>=4.0.0->ipywidgets) (2.14.0)
Requirement already satisfied: pickleshare in /usr/local/lib/python3.8/dist-packages (from ipython>=4.0.0->ipywidgets) (0.7.5)
Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.8/dist-packages (from widgetsnbextension<=3.6.0->ipywidgets) (6.3.0)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in /usr/local/lib/python3.8/dist-packages (from jedi>=0.10->ipython>=4.0.0->ipywidgets) (0.8.3)
Requirement already satisfied: nbconvert in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (6.1.0)
Requirement already satisfied: pyzmq>=17 in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (23.2.0)
Requirement already satisfied: argon2-cffi in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (21.1.0)
Requirement already satisfied: jupyter-core>=4.6.1 in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (4.7.0)
Requirement already satisfied: nbformat in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (5.1.0)
Requirement already satisfied: Send2Trash>=1.5.0 in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (1.8.0)
Requirement already satisfied: prometheus-client in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (0.12.1)
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (3.1.2)
Requirement already satisfied: terminado>=0.8.3 in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (0.15.0)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (2.8.1)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.8/dist-packages (from prompt-toolkit<2.1.0,>=2.0.0->ipython>=4.0.0->ipywidgets) (1.16.0)
Requirement already satisfied: wcwidth in /usr/local/lib/python3.8/dist-packages (from prompt-toolkit<2.1.0,>=2.0.0->ipython>=4.0.0->ipywidgets) (0.2.5)
Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.8/dist-packages (from pexpect->ipython>=4.0.0->ipywidgets) (0.7.0)
Requirement already satisfied: platformdirs>=2.5 in /usr/local/lib/python3.8/dist-packages (from jupyter-core>=4.6.1->notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (3.10.0)
Requirement already satisfied: argon2-cffi-bindings in /usr/local/lib/python3.8/dist-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (21.1.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.8/dist-packages (from Jinja2->notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (2.1.1)
Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.8/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (1.5.1)
Requirement already satisfied: bleach in /usr/local/lib/python3.8/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (4.1.0)
Requirement already satisfied: defusedxml in /usr/local/lib/python3.8/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (0.7.0)
Requirement already satisfied: BeautifulSoup4 in /usr/local/lib/python3.8/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (4.12.2)
Requirement already satisfied: packaging in /usr/local/lib/python3.8/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (23.1)
Requirement already satisfied: lxml in /usr/local/lib/python3.8/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension<=3.6.0->ipywidgets) (4.9.1)
```

```
Requirement already satisfied: noclnt>=0.5.0 in /usr/local/lib/python3.8/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.8/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.8/dist-packages (from nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: jsonschema>=2.6 in /usr/local/lib/python3.8/dist-packages (from nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: fastjsonschema in /usr/local/lib/python3.8/dist-packages (from nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: importlib-resources>=1.4.0 in /usr/local/lib/python3.8/dist-packages (from jsonschema>=2.6->nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: pyrsistent!=0.17.0,!0.17.1,!0.17.2,>=0.14.0 in /usr/local/lib/python3.8/dist-packages (from jsonschema>=2.6->nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.8/dist-packages (from jsonschema>=2.6->nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: cffi>=1.0.1 in /usr/local/lib/python3.8/dist-packages (from argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.8/dist-packages (from bleach->nbconvert->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: pycparser in /usr/local/lib/python3.8/dist-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.6.0)
Requirement already satisfied: zipp>=3.1.0 in /usr/local/lib/python3.8/dist-packages (from importlib-resources>=1.4.0->jsonschema>=2.6->nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0)
```

Reading a file

There is a dataset provided to everyone (in TSV format). Use the `read_file` API to read the file, store it in a variable simultaneously and print it.

Note: this API requires the path of your dataset as a required argument and `filetype` as an optional argument. API doc: https://psm-utils.readthedocs.io/en/v0.2.1/api/psm_utils.io/#psm_utils.io.read_file

```
psm_utils.io.read_file(filename: str | Path, *args, filetype: str = 'infer', **kwargs)
```

Read PSM file into `PSMList`.

Parameters:: `filename` (*str*) – Path to file.

`filetype` (*str, optional*) – File type. Any PSM file type with read support. See `psm_utils` tag in [Supported file formats](#).

`*args` (*tuple*) – Additional arguments are passed to the `psm_utils.io` reader.

`**kwargs` (*dict, optional*) – Additional keyword arguments are passed to the `psm_utils.io` reader.

```
#Write your code in this cell
file = read_file("/content/data.tsv")
print(file)
```

```
spectrum=None,
is_decoy=None,
```

```

        rank=None,
        source=None,
        provenance_data={},
        metadata={},
        rescoring_features={}
    ),
    ... +59
]
)

```

▼ File conversion

Now, use "io.convert" API to convert the input file to any other possible format (look at the documentation home page). This API requires the input file's path and the path where the output file should be stored as mandatory arguments. Set "show_progressbar = True" as an additional optional argument if you want to. API doc: https://psm-utils.readthedocs.io/en/v0.2.1/api/psm_utils.io/#psm_utils.io.convert

```

psm_utils.io.convert(input_filename: str | Path, output_filename: str | Path, input_filetype: str = 'infer',
output_filetype: str = 'infer', show_progressbar: bool = False)

```

Convert a PSM file from one format into another.

Parameters::

- input_filename** (*str*) – Path to input file.
- output_filename** (*str*) – Path to output file.
- input_filetype** (*str, optional*) – File type. Any PSM file type with read support. See psm_utils tag in [Supported file formats](#).
- output_filetype** (*str, optional*) – File type. Any PSM file type with write support. See psm_utils tag in [Supported file formats](#).
- show_progressbar** (*bool, optional*) – Show progress bar for conversion process. (default: False)

```

from pickle import TRUE
#Write your code in this cell
from psm_utils.io import convert
convert("/content/data.tsv", "/content/output_data.peprec", input_filetype = "tsv", output_filetype = "peprec", show_progressbar = TRUE)

```

Converting File

▼ Other APIs

Use any other API of your choice and play around with the tool.

```

from psm_utils.io.tsv import TSVReader
#Write your code in this cell
file_tsv = TSVReader("/content/data.tsv")
print(file_tsv)

#there weren't a lot of options for printing the data in the data file we were given
#I'd love to touch on this more in class!

<psm_utils.io.tsv.TSVReader object at 0x7f28e84b6520>

```

Submissions

Attach your code and submit it in assignment 2.

