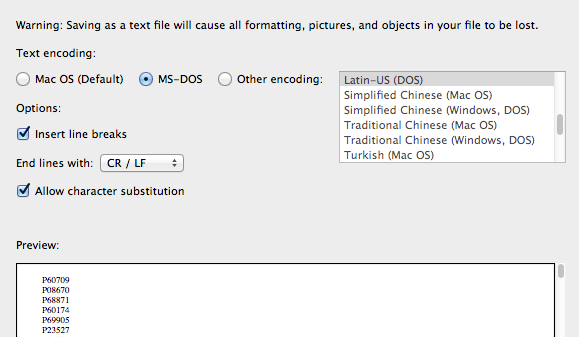
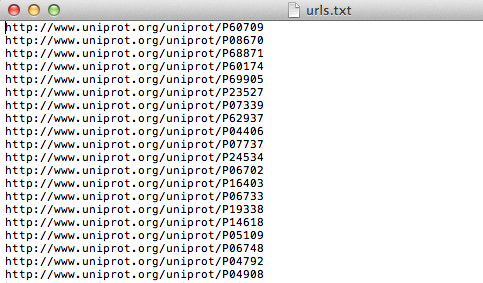
**Creating Text File of UniProt URLs**

1. From the Excel file with the list of accession numbers, highlight the column (“cmd-shift-downarrow”) of non-duplicated accession numbers grouped by decreasing number of peptides per protein. Copy and paste that into a Word document.
2. Delete column title “Accession number.”
3. Scroll all the way down to the last accession number. Notice that there may be many empty rows after that. If there are, add a page break after the very last accession by going to Layout 🡪 Break 🡪 Page. The rest of the column will show up on the following page. Go to that page, click on the 4-way arrow drag icon and delete it.
4. Save document as accessions.txt (change file name according to what you want) and check “Insert line breaks”, “End lines with ‘CR/LF’”, and “Allow character substitution”. Go to text file and delete any empty space in the beginning of the file and any extra UniProt URLs without accession numbers.



1. Go on Python script **accession\_numbers\_into\_uniprot\_url.py** by going into Desktop 🡪 PycharmProjects 🡪 PDB. Right click on the Python file and open with “IDLE 2.7”. \*\*\*Make sure it is Python 2.7, because this code was written for that and not for Python 3.5.
2. Macintosh HD:Users:legalllab:Desktop:Screen Shot 2017-05-23 at 3.09.17 PM.pngOn line 5, change name and path of file ‘accessions.txt’ with the name of the text file with the accession numbers you just created in the previous steps.
3. Macintosh HD:Users:legalllab:Desktop:Screen Shot 2017-05-23 at 3.09.30 PM.pngOn line 13, change ‘urls.txt’ to a name that you want the new text file of UniProt URLs to be. You can save it in a specific place on your device is you specify the path.
4. Run script with F5 or click Run 🡪 Run Module on the taskbar.
5. A file with the name you just replaced ‘urls.txt’ with will be created in the Documents folder. You can save it in a specific place on your device is you specify the path.

Your text file should look like the following:



**Querying UniProt URLs for Protein Information**

1. Go on Python script **protein\_name\_sequence\_multiple\_uniprot\_urls.py**.
2. On line 78, change name of file ‘uniprot\_url\_txt’ with the path and name of the text file with the accession numbers you just created in the previous steps.

Macintosh HD:Users:legalllab:Desktop:Screen Shot 2018-08-07 at 1.46.39 PM.png

1. Save this Python script so that later when you run Terminal, it reads an updated version to parse through the correct text file with all the URLs.
2. (OPTIONAL STEP) If you want to check and see that the script is reading the correct file of URLs, you can run with F5 or Run 🡪 Run Module on the taskbar. After the first couple of proteins run successfully, you can kill the program.

**Transferring Python Script Information Onto Excel File**

1. On Terminal, change path to the location of **protein\_name\_sequence\_multiple\_uniprot\_urls.py** like the following:

$ cd $HOME/PycharmProjects

1. Type into the command line the following where you can change the name of your Excel file to a name you think is appropriate to easily identify later in your files.

$ python PDB/protein\_name\_sequence\_multiple\_uniprot\_urls.py > PDBResults/protinfo\_name\_date.xls

1. You may need to wait a while (and by while I mean like about 35 minutes for 1500 peptides so the more peptides, the longer amount of time unfortunately) for this to run (until Terminal requests for another command denoted by $) before opening the Excel file with all the secondary structure information.
2. On the Excel file, separate the information into separate columns by selecting the column and going to Data🡪Text to Columns 🡪 check Delimited 🡪 check Space and Treat consecutive delimiters as one 🡪 Next 🡪 Finish.

