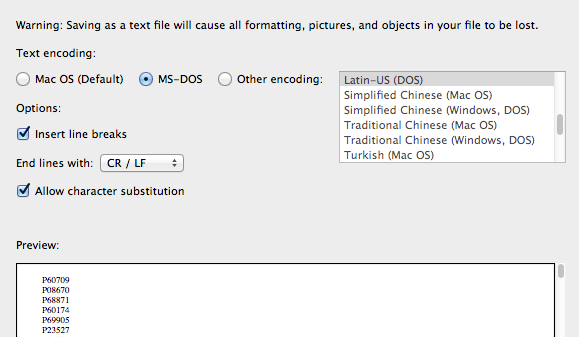
**Creating Text File of UniProt URLs**

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.

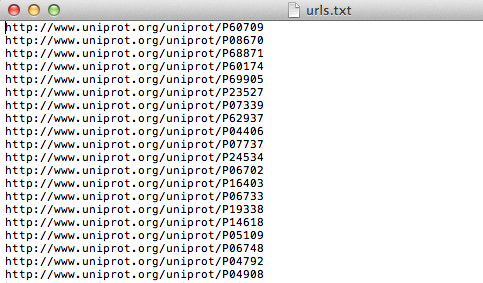


1. Go on Python script **accession\_numbers\_into\_uniprot\_url.py**.
2. Macintosh HD:Users:legalllab:Desktop:Screen Shot 2017-05-23 at 3.09.17 PM.pngOn line 5, change name and path of file ‘accessions.txt’ with the name of the text file with the accession numbers you just created in the previous steps.
3. On 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.

Macintosh HD:Users:legalllab:Desktop:Screen Shot 2017-05-23 at 3.09.30 PM.png

1. Run script with F5 or click Run 🡪 Run Module on the taskbar.
2. 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 Secondary Structure Information**

1. Go on Python script **secondary\_structure\_through\_multiple\_uniprot\_urls.py**.
2. On line 5, change name of file ‘accessions.txt’ with the name of the text file with the accession numbers you just created in the previous steps.

Macintosh HD:Users:legalllab:Desktop:Screen Shot 2017-05-23 at 3.36.24 PM.png

1. (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**

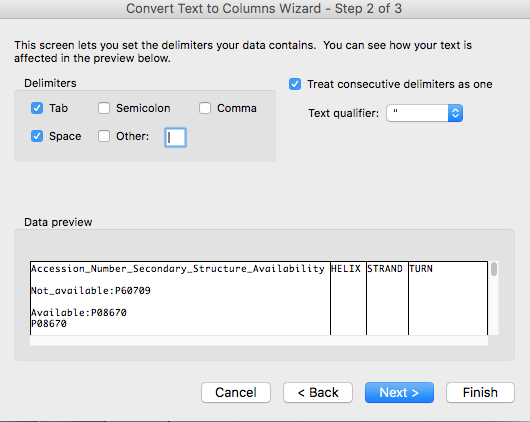
1. On Terminal, change path to the location of **secondary\_structure\_through\_multiple\_uniprot\_urls.py** like the following:

$ cd $HOME/PycharmProjects

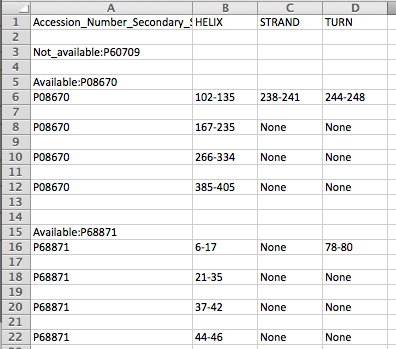
1. Type into the command line the following where you can change the name of your Excel file to a name you think is appropriate to easily identify later in your files.

$ python PDB/secondary\_structure\_through\_multiple\_uniprot\_urls.py > PDBResults/secstruc\_name\_date.xls

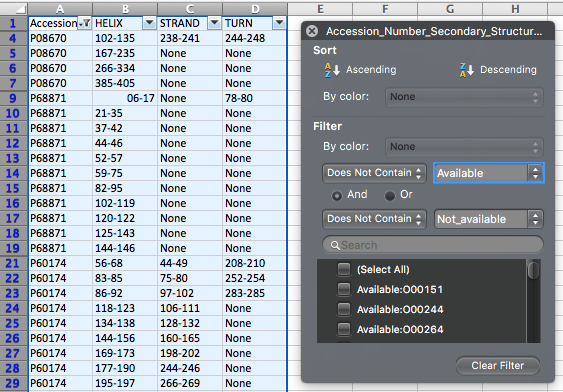
1. You may need to wait a while (and by while I mean like about 35 minutes for 1500 peptides so the more peptides, the longer amount of time unfortunately) for this to run (until Terminal requests for another command denoted by $) before opening the Excel file with all the secondary structure information.
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.



1. To reformat the columns, delete one empty row from rows B through D in order to shift the cells up.
2. Highlight all of Column C, D, and E, right-click the mouse and select “Format Cells…” Under “Number”, choose “Custom” and type in “mm-dd”. The Excel sheet should look something like the following right now.



1. Highlight columns A through D, click the “Filter” button under the Data section, and filter by checking “Select All” first, then scrolling to the bottom of this list and unchecking “Blanks” and “Successful!”
2. Open a new sheet and copy and paste columns A-D. Highlight columns A through D, click the “Filter” button under the Data section, and filter with “Does Not Contain: Available” and “Does not contain: Not\_available” like the following.



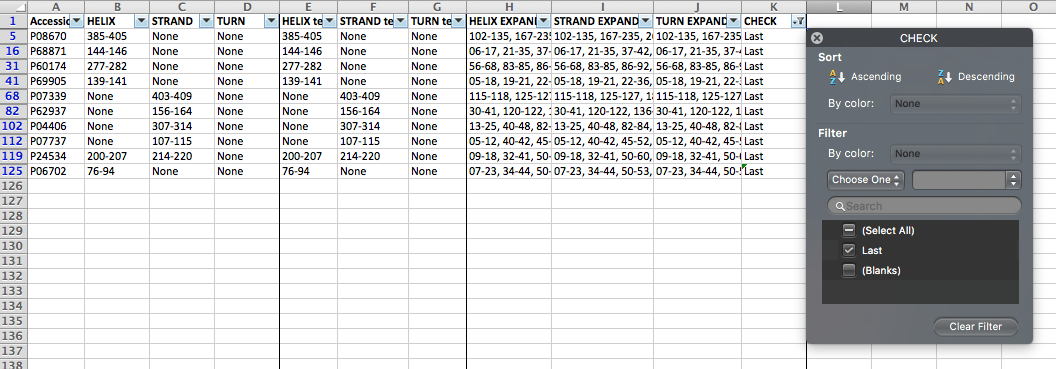
**Using “Matching Secondary Structure.xlsm”**

1. Open the Excel sheet named “Matching Secondary Structure” and copy and paste columns A-D into the sheet named “edit\_output”.
2. In the following cells, these formulas are already entered in so just highlight across B2-K2 and double-click on the right-bottom corner of the cell to copy the formula down the column.

**E2**: =TEXT(B2,"mm-dd"), **F2**: =TEXT(C2,"mm-dd"), **G2**: =TEXT(D2,"mm-dd"), **H2**: =IF(A2=A1,H1&", "&E2, E2)

**I2:** =IF(A2=A1,I1&", "&F2, F2), **J2**: =IF(A2=A1,J1&", "&G2, G2), **K2**: =IF(A2<>A3, "Last", "")

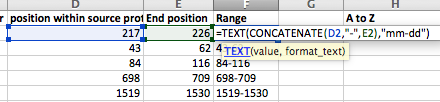
1. Highlight all columns. Under the “Data” tab, select “Filter”. Under the “CHECK” or K column, uncheck “Blanks” in the drop down menu.



1. Highlight columns H through K, press “command-F” and replace all “None” with blanks. Copy columns A(Accession\_Number\_Secondary\_Structure\_Availability), H(HELIX RANGE), I(STRAND RANGE), and J(TURN RANGE) into columns A-D in the “ref” sheet. Select entire columns B through D.

*Useful trick: Whenever you want to select information down a column until the very last cell with information inside, click on the column title or drag across the column titles you want to copy and then press “Command-shift-down arrow”.*

1. Macintosh HD:Users:legalllab:Desktop:Screen Shot 2017-07-14 at 12.04.43 PM.pngFrom the Excel sheet with all the given data, copy the “Peptide sequences” and “Accession number” information into the columns E and F. In that same Excel sheet, make a column named “End position” and type in a formula that adds numbers in the columns “Position within source prot” and “Peptide length” such as the following.
2. Make another column named “Range” and type in a formula that concatenates the numbers in the columns “Position within source prot” and “End position” such as the following. Copy and paste special as “Values only” the range into columns G and H of the “ref” sheet in the “Matching Secondary Structure” Excel sheet.



1. *(This step is already completed for the user but this is to explain where the formula comes from)*

On the sheet, “ref”, click the button labeled “Find H,S,T”. This is what happens:

Under the Developer tab, open Editor. Right click “Sheet 6(ref)”🡪 Insert 🡪 Module. Copy and paste the following VBA code.

**Function NumRange(v)**

**Dim arrC, arr, x As Long, rv As String, sep As String, e**

**arrC = Split(v, ",")**

**rv = ""**

**For Each e In arrC**

**If InStr(e, "-") Then**

**arr = Split(e, "-")**

**arr(0) = Trim(arr(0))**

**arr(1) = Trim(arr(1))**

**If IsNumeric(arr(0)) And IsNumeric(arr(1)) Then**

**For x = CLng(arr(0)) To CLng(arr(1))**

**rv = rv & sep & x**

**sep = ","**

**Next x**

**End If**

**ElseIf IsNumeric(e) Then**

**rv = rv & sep & CLng(e)**

**sep = ","**

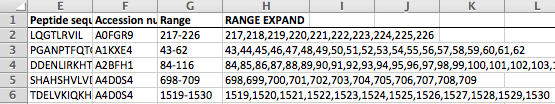
**End If**

**Next e**

**NumRange = rv**

**End Function**

1. In the “RANGE EXPAND” column, type in “NumRange(cell)”. For cell inside the formula, select the adjacent cell under the “Range” column. The formula will already be entered in H2, so just double-click the bottom-right corner of the cell. The result will turn out to look like the following:



In the following cells, these formulas are already entered in so just highlight I2-N2 and double-click on the right-bottom corner of the last cell to copy the formula down the columns.

**I2 (Find Helix)**: =IFERROR(VLOOKUP(F2,$A:$D,2,FALSE),""), **J2 (HELIX EXPAND)**: =NumRange(I2)

**K2** **(Find Strand**): =IFERROR(VLOOKUP(F2,$A:$D,3,FALSE),""), **L2 (STRAND EXPAND)**: =NumRange(K2)

**M2 (Find Turn)**: =IFERROR(VLOOKUP(F2,$A:$D,4,FALSE),""), **N2 (TURN EXPAND)**: =NumRange(M2)

1. On the sheet, “range”, click on “Copy Peptides and Expand” which copies the peptide sequences accession numbers, and column “RANGE EXPAND” from the “ref” sheet.
2. On the sheet, “ss”, click on “Copy Columns A and B” which copies the peptide sequences accession numbers into the “ss”, “match” and “results” sheet.
3. On the sheet, “ss”, click on “Copy H” which copies the column “HELIX EXPAND” from the “ref” sheet. It will expand the helix ranges so that each number is in a separate cell across rows. When you come back to this step for strands and turns, click on “Clear H,S,T” to delete the secondary information and replace by clicking on “Copy S” or “Copy T”, depending on which structure you are concentrating on.
4. On the sheet, “results”, the following formula is in C7 (H, S or T). Double-click the bottom-right corner of the cell to apply the formula down the column. Copy and paste the “H,S,or T” column into the corresponding structure column.

=match!C7&IF(match!D7<>"",","&match!D7,"")&IF(match!E7<>"",","&match!E7,"")&IF(match!F7<>"",","&match!F7,"")&IF(match!G7<>"",","&match!G7,"")&IF(match!H7<>"",","&match!H7,"")&IF(match!I7<>"",","&match!I7,"")&IF(match!J7<>"",","&match!J7,"")&IF(match!K7<>"",","&match!K7,"")&IF(match!L7<>"",","&match!L7,"")&IF(match!M7<>"",","&match!M7,"")&IF(match!N7<>"",","&match!N7,"")&IF(match!O7<>"",","&match!O7,"")&IF(match!P7<>"",","&match!P7,"")&IF(match!Q7<>"",","&match!Q7,"")&IF(match!R7<>"",","&match!R7,"")&IF(match!S7<>"",","&match!S7,"")&IF(match!T7<>"",","&match!T7,"")&IF(match!U7<>"",","&match!U7,"")&IF(match!V7<>"",","&match!V7,"")&IF(match!W7<>"",","&match!W7,"")&IF(match!X7<>"",","&match!X7,"")&IF(match!Y7<>"",","&match!Y7,"")&IF(match!Z7<>"",","&match!Z7,"")&IF(match!AA7<>"",","&match!AA7,"")&IF(match!AB7<>"",","&match!AB7,"")&IF(match!AC7<>"",","&match!AC7,"")&IF(match!AD7<>"",","&match!AD7,"")&IF(match!AE7<>"",","&match!AE7,"")&IF(match!AF7<>"",","&match!AF7,"")&IF(match!AG7<>"",","&match!AG7,"")&IF(match!AH7<>"",","&match!AH7,"")&IF(match!AI7<>"",","&match!AI7,"")&IF(match!AJ7<>"",","&match!AJ7,"")&IF(match!AK7<>"",","&match!AK7,"")&IF(match!AL7<>"",","&match!AL7,"")&IF(match!AM7<>"",","&match!AM7,"")&IF(match!AN7<>"",","&match!AN7,"")&IF(match!AO7<>"",","&match!AO7,"")&IF(match!AP7<>"",","&match!AP7,"")&IF(match!AQ7<>"",","&match!AQ7,"")&IF(match!AR7<>"",","&match!AR7,"")&IF(match!AS7<>"",","&match!AS7,"")&IF(match!AT7<>"",","&match!AT7,"")&IF(match!AU7<>"",","&match!AU7,"")&IF(match!AV7<>"",","&match!AV7,"")&IF(match!AW7<>"",","&match!AW7,"")&IF(match!AX7<>"",","&match!AX7,"")&IF(match!AY7<>"",","&match!AY7,"")&IF(match!AZ7<>"",","&match!AZ7,"")&IF(match!BA7<>"",","&match!BA7,"")&IF(match!BB7<>"",","&match!BB7,"")&IF(match!BC7<>"",","&match!BC7,"")&IF(match!BD7<>"",","&match!BD7,"")&IF(match!BE7<>"",","&match!BE7,"")&IF(match!BF7<>"",","&match!BF7,"")&IF(match!BG7<>"",","&match!BG7,"")&IF(match!BH7<>"",","&match!BH7,"")&IF(match!BI7<>"",","&match!BI7,"")&IF(match!BJ7<>"",","&match!BJ7,"")&IF(match!BK7<>"",","&match!BK7,"")&IF(match!BL7<>"",","&match!BL7,"")&IF(match!BM7<>"",","&match!BM7,"")&IF(match!BN7<>"",","&match!BN7,"")&IF(match!BO7<>"",","&match!BO7,"")&IF(match!BP7<>"",","&match!BP7,"")&IF(match!BQ7<>"",","&match!BQ7,"")&IF(match!BR7<>"",","&match!BR7,"")&IF(match!BS7<>"",","&match!BS7,"")&IF(match!BT7<>"",","&match!BT7,"")&IF(match!BU7<>"",","&match!BU7,"")&IF(match!BV7<>"",","&match!BV7,"")&IF(match!BW7<>"",","&match!BW7,"")&IF(match!BX7<>"",","&match!BX7,"")&IF(match!BY7<>"",","&match!BY7,"")&IF(match!BZ7<>"",","&match!BZ7,"")

*Note: DO NOT DELETE OR ALTER ANY CELL FORMULAS FROM THE “MATCH” OR “RESULTS” SHEET.*

These are the formulas in the “match” sheet for reference and future applications.

=IF(COUNTIF(ss!7:7,range!AA7)=0, "",range!AA7)

=IF(COUNTIF(ss!7:7,range!AB7)=0, "",range!AB7)

=IF(COUNTIF(ss!7:7,range!AC7)=0, "",range!AC7)

=IF(COUNTIF(ss!7:7,range!AD7)=0, "", range!AD7)

=IF(COUNTIF(ss!7:7,range!AE7)=0, "", range!AE7)

=IF(COUNTIF(ss!7:7,range!AF7)=0, "", range!AF7)

=IF(COUNTIF(ss!7:7,range!AG7)=0, "", range!AG7)

=IF(COUNTIF(ss!7:7,range!AH7)=0, "", range!AH7)

=IF(COUNTIF(ss!7:7,range!AI7)=0, "", range!AI7)

=IF(COUNTIF(ss!7:7,range!AJ7)=0, "", range!AJ7)

=IF(COUNTIF(ss!7:7,range!AK7)=0, "", range!AK7)

=IF(COUNTIF(ss!7:7,range!AL7)=0, "", range!AL7)

=IF(COUNTIF(ss!7:7,range!AM7)=0, "", range!AM7)

=IF(COUNTIF(ss!7:7,range!AN7)=0, "", range!AN7)

=IF(COUNTIF(ss!7:7,range!AO7)=0, "", range!AO7)

=IF(COUNTIF(ss!7:7,range!AP7)=0, "", range!AP7)

=IF(COUNTIF(ss!7:7,range!AQ7)=0, "", range!AQ7)

=IF(COUNTIF(ss!7:7,range!AR7)=0, "", range!AR7)

=IF(COUNTIF(ss!7:7,range!AS7)=0, "", range!AS7)

=IF(COUNTIF(ss!7:7,range!AT7)=0, "", range!AT7)

=IF(COUNTIF(ss!7:7,range!AU7)=0, "", range!AU7)

=IF(COUNTIF(ss!7:7,range!AV7)=0, "", range!AV7)

=IF(COUNTIF(ss!7:7,range!AW7)=0, "", range!AW7)

=IF(COUNTIF(ss!7:7,range!AX7)=0, "", range!AX7)

=IF(COUNTIF(ss!7:7,range!AY7)=0, "", range!AY7)

=IF(COUNTIF(ss!7:7,range!AZ7)=0, "", range!AZ7)

=IF(COUNTIF(ss!7:7,range!BA7)=0, "",range!BA7)

=IF(COUNTIF(ss!7:7,range!BB7)=0, "",range!BB7)

=IF(COUNTIF(ss!7:7,range!BC7)=0, "",range!BC7)

=IF(COUNTIF(ss!7:7,range!BD7)=0, "", range!BD7)

=IF(COUNTIF(ss!7:7,range!BE7)=0, "", range!BE7)

=IF(COUNTIF(ss!7:7,range!BF7)=0, "", range!BF7)

=IF(COUNTIF(ss!7:7,range!BG7)=0, "", range!BG7)

=IF(COUNTIF(ss!7:7,range!BH7)=0, "", range!BH7)

=IF(COUNTIF(ss!7:7,range!BI7)=0, "", range!BI7)

=IF(COUNTIF(ss!7:7,range!BJ7)=0, "", range!BJ7)

=IF(COUNTIF(ss!7:7,range!BK7)=0, "", range!BK7)

=IF(COUNTIF(ss!7:7,range!BL7)=0, "", range!BL7)

=IF(COUNTIF(ss!7:7,range!BM7)=0, "", range!BM7)

=IF(COUNTIF(ss!7:7,range!BN7)=0, "", range!BN7)

=IF(COUNTIF(ss!7:7,range!BO7)=0, "", range!BO7)

=IF(COUNTIF(ss!7:7,range!BP7)=0, "", range!BP7)

=IF(COUNTIF(ss!7:7,range!BQ7)=0, "", range!BQ7)

=IF(COUNTIF(ss!7:7,range!BR7)=0, "", range!BR7)

=IF(COUNTIF(ss!7:7,range!BS7)=0, "", range!BS7)

=IF(COUNTIF(ss!7:7,range!BT7)=0, "", range!BT7)

=IF(COUNTIF(ss!7:7,range!BU7)=0, "", range!BU7)

=IF(COUNTIF(ss!7:7,range!BV7)=0, "", range!BV7)

=IF(COUNTIF(ss!7:7,range!BW7)=0, "", range!BW7)

=IF(COUNTIF(ss!7:7,range!BX7)=0, "", range!BX7)

=IF(COUNTIF(ss!7:7,range!BY7)=0, "", range!BY7)

=IF(COUNTIF(ss!7:7,range!BZ7)=0, "", range!BZ7)

=IF(COUNTIF(ss!7:7,range!C7)=0, "",range!C7)

=IF(COUNTIF(ss!7:7,range!D7)=0, "", range!D7)

=IF(COUNTIF(ss!7:7,range!E7)=0, "", range!E7)

=IF(COUNTIF(ss!7:7,range!F7)=0, "", range!F7)

=IF(COUNTIF(ss!7:7,range!G7)=0, "", range!G7)

=IF(COUNTIF(ss!7:7,range!H7)=0, "", range!H7)

=IF(COUNTIF(ss!7:7,range!I7)=0, "", range!I7)

=IF(COUNTIF(ss!7:7,range!J7)=0, "", range!J7)

=IF(COUNTIF(ss!7:7,range!K7)=0, "", range!K7)

=IF(COUNTIF(ss!7:7,range!L7)=0, "", range!L7)

=IF(COUNTIF(ss!7:7,range!M7)=0, "", range!M7)

=IF(COUNTIF(ss!7:7,range!N7)=0, "", range!N7)

=IF(COUNTIF(ss!7:7,range!O7)=0, "", range!O7)

=IF(COUNTIF(ss!7:7,range!P7)=0, "", range!P7)

=IF(COUNTIF(ss!7:7,range!Q7)=0, "", range!Q7)

=IF(COUNTIF(ss!7:7,range!R7)=0, "", range!R7)

=IF(COUNTIF(ss!7:7,range!S7)=0, "", range!S7)

=IF(COUNTIF(ss!7:7,range!T7)=0, "", range!T7)

=IF(COUNTIF(ss!7:7,range!U7)=0, "", range!U7)

=IF(COUNTIF(ss!7:7,range!V7)=0, "", range!V7)

=IF(COUNTIF(ss!7:7,range!W7)=0, "", range!W7)

=IF(COUNTIF(ss!7:7,range!X7)=0, "", range!X7)

=IF(COUNTIF(ss!7:7,range!Y7)=0, "", range!Y7)

=IF(COUNTIF(ss!7:7,range!Z7)=0, "", range!Z7)

1. Follows steps 10-12 for “Strands” and “Turns”. For a repeat for strands and turns, you do not need to change any formulas for step 12. All you need to do it copy and paste the column into the corresponding “Strands” or “Turns” column in the “results” sheet.
2. Voila! You are all done! C’est bien!