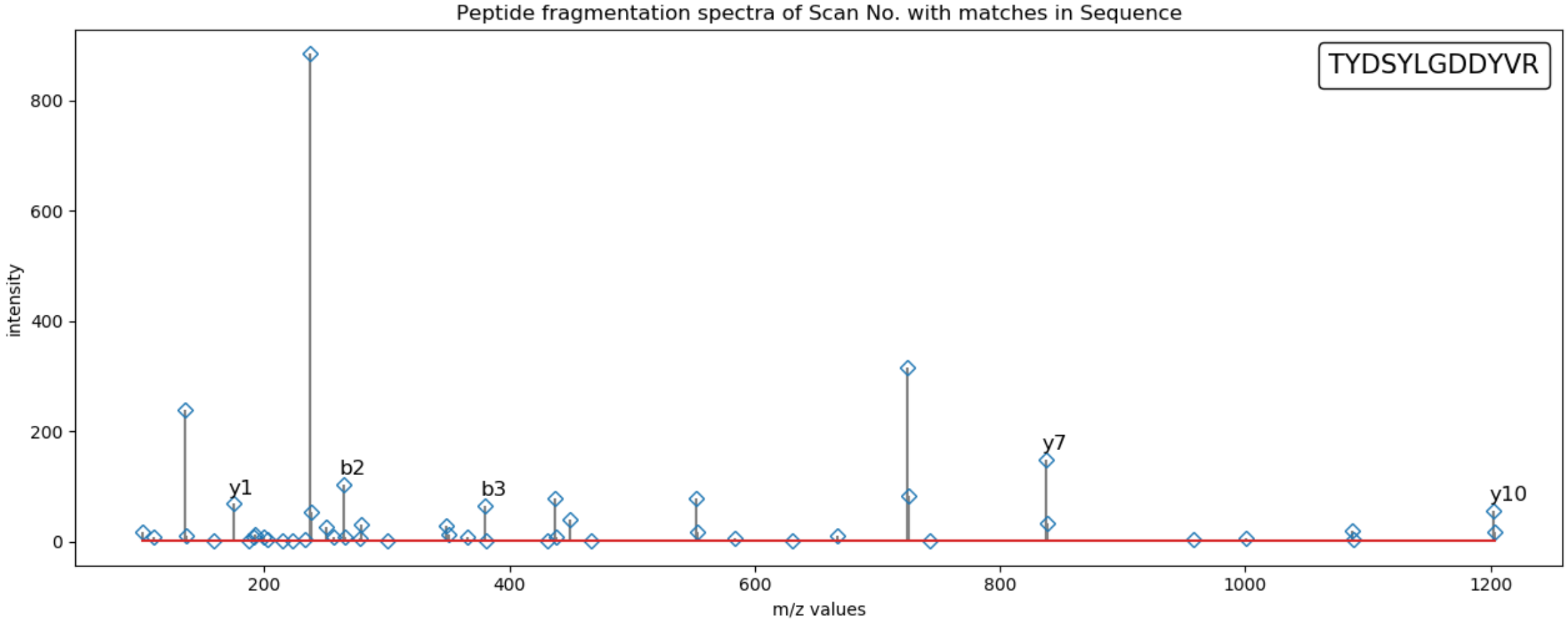
MS/MS VIEWER

Helen Ashdown

Objective: Display peptide fragmentation spectra from an mzXMLfile

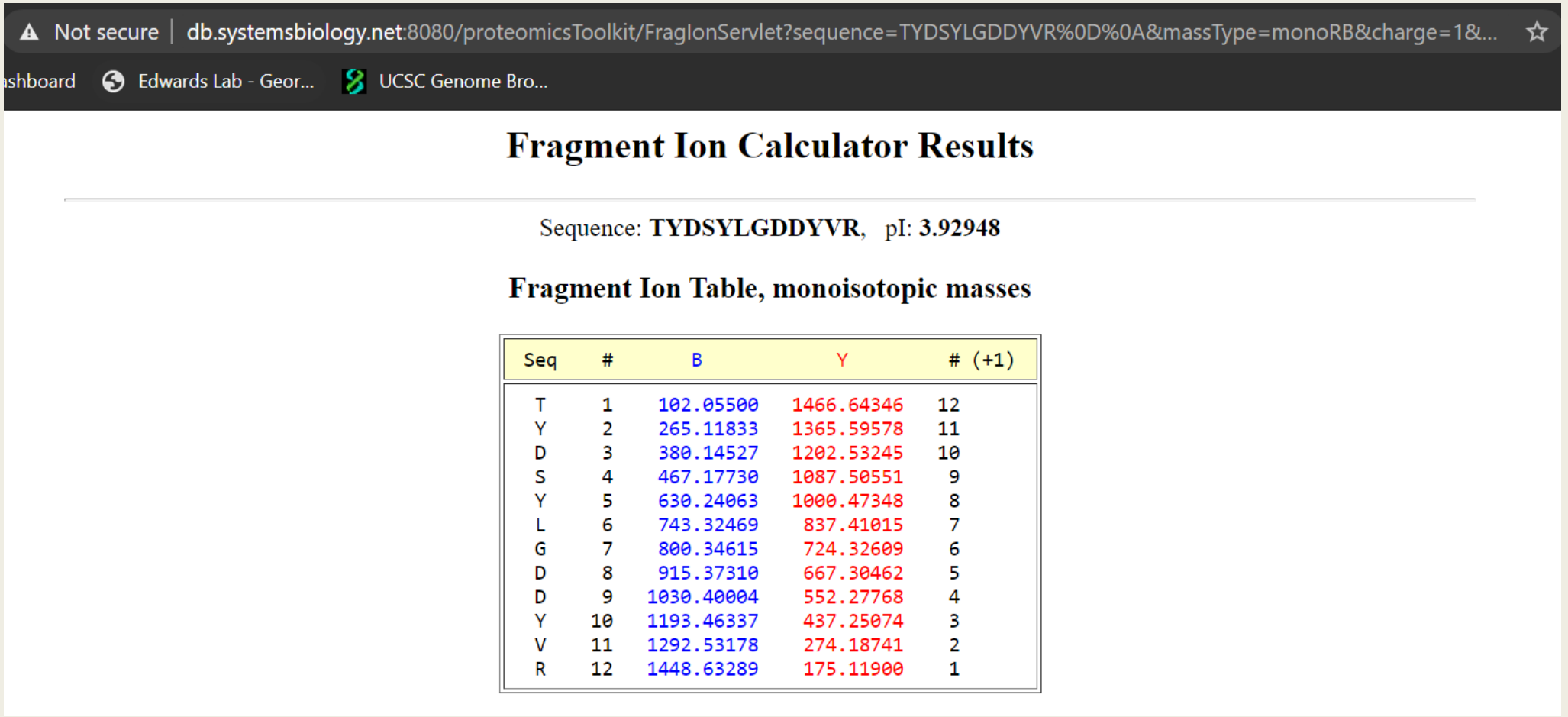
- Input:
  - *mzXML file:* 17mix_test2.mzxml.gz
  - Scan number: 1298
  - Peptide sequence: TYDSYLGDDYVR
- Requirements:
  - Extract given spectra from file using scan number
  - Peptide’s b-ion and y-ion m/z values computed (Theoretical)
  - Annotate spectra peaks which match the theoretical values
  - Use output figure to determine whether peptide is a good match to spectrum


Peptide fragmentation spectra of Scan No. with matches in Sequence
TYDSYLGDDYVR
intensity
m/z values
y1
b2
b3
y7
y10
0
200
400
600
800
200
400
600
800
1000
1200

Trade-offs

- I should have used a better data structure for my matches to eventually pull out the relevant information needed to plot them on my spectra graph.
- Keeping track of the matches data set to plot could be more concise.
- Possible other routes for data structure would be an overall encompassing function for matches that include all

How I tested calculated theoretical B and Y ion values:



Fragment Ion Calculator Results

Sequence: **TYDSYLGDDYVR**, pI: **3.92948**

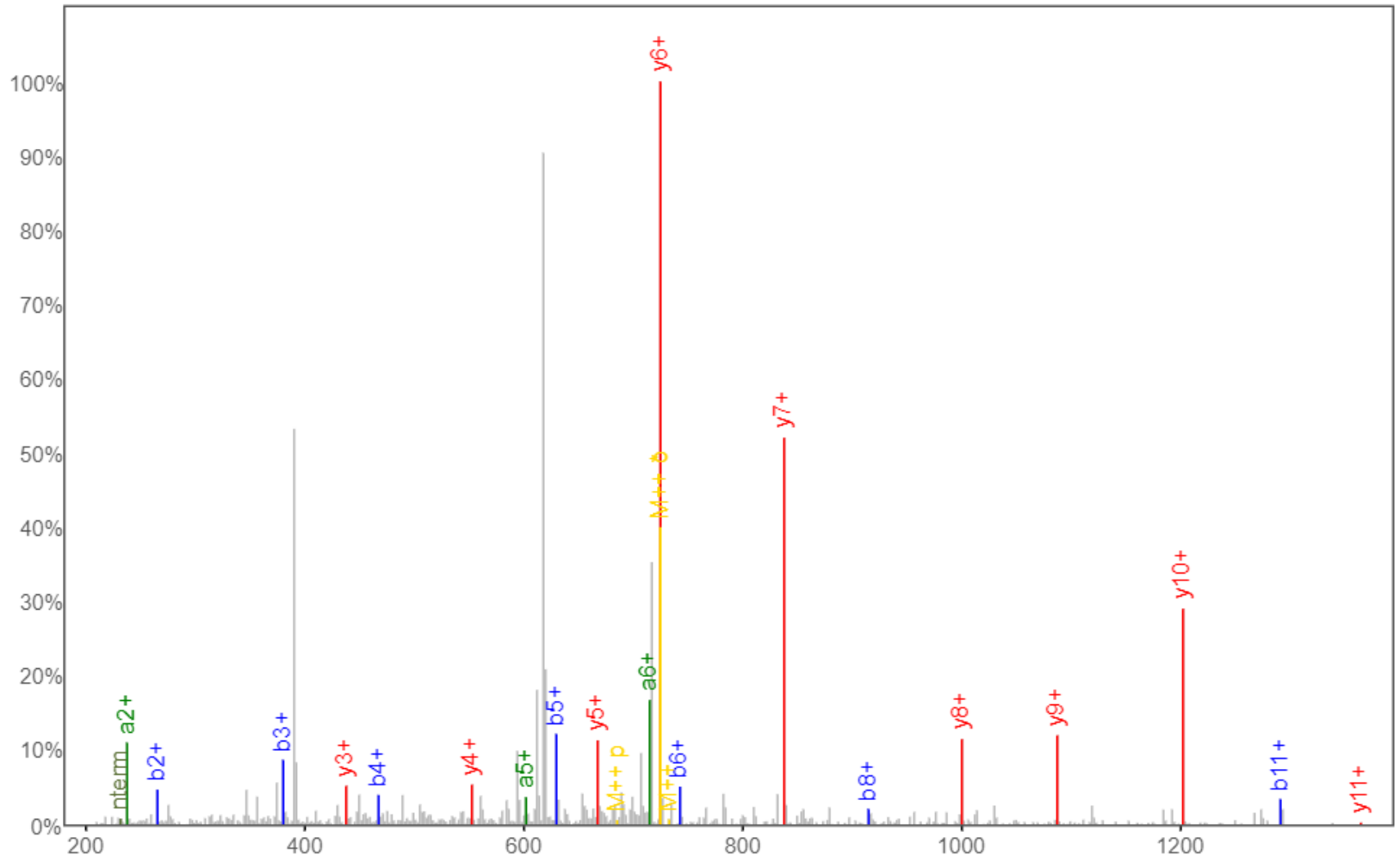
Fragment Ion Table, monoisotopic masses

Seq	#	B	Y	# (+1)
T	1	102.05500	1466.64346	12
Y	2	265.11833	1365.59578	11
D	3	380.14527	1202.53245	10
S	4	467.17730	1087.50551	9
Y	5	630.24063	1000.47348	8
L	6	743.32469	837.41015	7
G	7	800.34615	724.32609	6
D	8	915.37310	667.30462	5
D	9	1030.40004	552.27768	4
Y	10	1193.46337	437.25074	3
V	11	1292.53178	274.18741	2
R	12	1448.63289	175.11900	1

(Less accurate) Found spectra of same peptide online

TYDSYLGDDYVR, MH+ 1466.6434, m/z 733.8253

File: 180924_OV1_FS__24_D2_Ust.05146.05146.2, Scan: 5146, Exp. m/z: 733.82533896688, Charge: 2



Click and drag in the plot to zoom X: ☑ Y: ☐ Zoom Out Print ☐ Enable tooltip ☑ Plot mass error

a+	b+	#	Seq	#	y+
74.0600	102.0550	1	T	12	
237.1234	**265.1183**	2	Y	11	**1365.5957**
352.1503	**380.1452**	3	D	10	**1202.5324**
439.1823	**467.1773**	4	S	9	**1087.5055**
602.2457	**630.2406**	5	Y	8	**1000.4734**
715.3297	**743.3246**	6	L	7	**837.4101**
772.3512	800.3461	7	G	6	**724.3260**
887.3781	**915.3731**	8	D	5	**667.3046**
1002.4051	1030.4000	9	D	4	**552.2776**
1165.4684	1193.4633	10	Y	3	**437.2507**
1264.5368	**1292.5317**	11	V	2	274.1874
		12	R	1	175.1190

[Click] to move table

- Tricks I learned:
  - *Simple math for the bion and yion m/z values*
  - *Using matplotlib to create scatter plot for spectra was relatively user friendly and easy to manipulate*
  - *Once both the sequence values and scan values were created and stored, executing the match criteria was relatively straight forward*
  - *Checking my own calculations from internet sources was quite helpful*

- Bugs I (also) learned:
  - *It took me quite awhile to even get to a point where I could even extract anything from xml file (used a lot of nested loops)*
  - *When I did get something, it was pure gibberish*
  - *Luckily Dr. Edwards provided me with magical code to extract my peaks element*
  - *Annotating matplotlib was not user friendly and I would not recommend using it if you want to write a program for say, wedding seat assignments (using text and figures was a drag)*

Modules and packages

- The most important package was the xml.etree.ElementTree as ET to parse through my xml file.
- Using the iterparse function was easily the most important component to get my project off the ground.
- Relatively user friendly and easily applicable to a lot of files and data we used this semester
- I have my own thoughts about the application of matplot lib to annotate my spectra, but overall using that package (along with pyplot and numpy) was very effective and I did like learning so much using matplotlib

- Aspects that were straight forward
  - *Assigning my x and y values for my spectra plot !*
  - *The fundamental conceptual goal of the project I found very easy to understand*
  - *Because of that, when I reached the point of having all of my data it was easier to plan my next steps*
  - *The workflow of this project was very straightforward too.*
  - *I used simple data structures because I knew they worked and accomplished the task I needed them too*

- Aspects that were straight backwards
  - *I began this project the week of Thanksgiving, and I am very glad that I did. Over the weeks, I would slowly add the next component of my workflow*
  - *Like Dr. Edwards had mentioned when he introduced this project, the longest and most arduous task was parsing the mzXMLfile.*
  - *Even after I was able to access its contents, it took me just as long to decipher what was going on (picking the correct element)*
  - *It took me an hour to add a label to my matplotlib spectra, also they have a few versions which was confusing too*

My big take aways

- During this project the B key fell off of my keyboard.
- What I learned about bioinformatics was how to analyze data in different forms and altering it to become comparable against data with the same ‘components’.
- i.e. taking the xml file and the peptide sequence and extracting the relevant data, then formatting that data in a way which made it be able to compare and contrast, thereby allowing us to draw conclusions about the information.
- Having no prior coding experience, I learned that programming is a lot like being a pianist like I am. It takes a lot of problem-solving *and* creativity, but I find myself cursing more at Beethoven than Dr. Edwards.
- All in all, I have come away from this semester with a set of skills applicable to many scientific and computational fields.
- And shoutout to Dr. Fauci.