Qualitative and Quantitative Analysis of Mass Spectrometry Peaks of Glycans

Shelly Heim, Anyin Li*
Chemistry, University of New Hampshire, Durham, NH 03824

Introduction

Measuring the correlation between various data sets is a method of assessing how strong the relationship between the data sets is and how they change simultaneously. It is essentially the process of finding a pattern between the data sets. Measuring correlation has many real-world applications. It is quite significant in medical fields, as it allows for connections between factors such as treatment and treatment outcome.[1] It is also a powerful tool in most science fields as a whole. Correlation can help to tell scientists why something happened, as it was likely due to another factor. The relationship between the outcome and various factors can be quantified and used to determine a better method to obtain the desired results for an experiment.[2]

Comparing results from different methods of analysis allows for the results to be verified. Using multiple methods to determine if the results are consistent across platforms is a key step in the data analysis process. After completing an experiment, the results can be compared to those found in literature to verify the numbers. Using two different programs to analyze the data and compare the results is another popular method of data verification. Precise data is extremely important in chemistry, thus meaning that the data analysis being correct is just as crucial to the understanding of chemistry.

Methodology

The Li research group used various instruments to collect data on procainamide labeled glucose homopolymer ladder. The un-analyzed data was split up amongst the group where it was then analyzed. The correlation of the data was calculated to determine the accuracy of the results from the various instruments used. Two methods of correlation analysis were used to do so: the Pearson Method and the analysis tool within Microsoft Excel (CORREL). All of the results from the analysis can be seen in tables 1-6.

A polymer structure, $C_{31}H_{53}N_3O_{16}$, was provided by Anyin Li to be analyzed using *ChemDraw* and *MatLab* programs. The structure was recreated using *ChemDraw* (figure 1), where the program provided the exact mass and the molecular weight of the structure (figure 2). Code was then written into *MatLab* that would calculate the mass of the polymer to then be compared with the results from *ChemDraw* (figure 3). The code written for *MatLab* was created with the help of an artificial intelligence code generator, *CodeConvert*. The A.I. code generator allowed for the creation of code that efficiently calculated the molecular mass of the given structure in the shortest amount of steps.

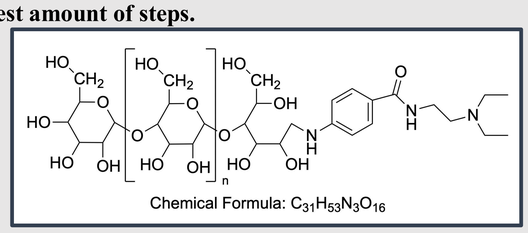

Figure 1

***MatLab* was also used by Professor Li to create a mass spectrometry spectrum for the given polymer structure (figure 4). The spectrum allowed for the confirmation of a peak at the determined mass of the polymer. The spectrum was analyzed by zooming into the 695-735$_{m/z}$ portion and finding the closest peak to the determined mass of the polymer (figure 5).**

Correlation Analysis

Table 1

LC-MS with nanoESI	LC-MS with femtoESI
0.795073094	0.923419648
nanoESI with femtoESI	
0.780605243	

Table 2

Method	LC-MS	nanoESI	femtoESI
LC-MS	1		
nanoESI	0.795073094	1	
femtoESI	0.923419648	0.780605243	1

Table 3

LC-MS with nanoESI	LC-MS with femtoESI
0.769831	0.960801
nanoESI with femtoESI	
0.742055	

Table 4

LCMS with nanoESI
0.841402564

Table 5

Method	LC-MS	nanoESI	femtoESI
LCMS	1		
nanoESI	0.769831	1	
femtoESI	0.960801	0.742055	1

Table 6

Method	LCMS	nanoESI
LCMS	1	
nanoESI	0.841403	1

Polymer Analysis

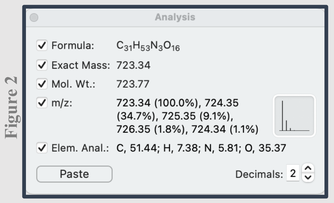

Figure 2

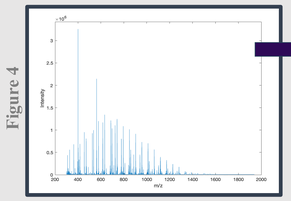

Figure 4

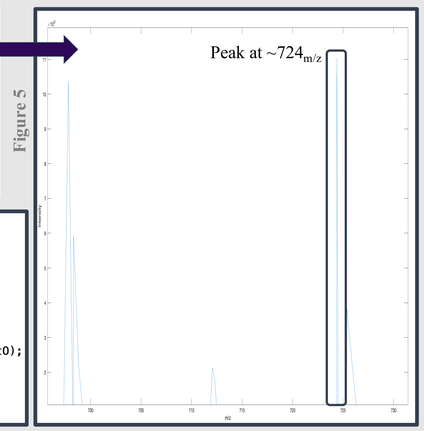

Figure 5

```
>> % Calculate the molecular weight of C31H54N3O16 with 4 decimal points accuracy

% Given atomic weights:
atomicWeightC = 12.00;   % Carbon
atomicWeightH = 1.008;   % Hydrogen
atomicWeightN = 14.00;   % Nitrogen
atomicWeightO = 15.99;   % Oxygen

% Calculate molecular weight directly using the formula:
% Molecular Weight = (31 * C) + (54 * H) + (3 * N) + (16 * O)
molecularWeight = (31 * atomicWeightC) + (54 * atomicWeightH) + (3 * atomicWeightN) + (16 * atomicWeightO);

% Display the result with 4 decimal point accuracy
fprintf('The molecular weight of C31H54N3O16 is %.4f\n', molecularWeight);
The molecular weight of C31H54N3O16 is 724.2720
>>
```

Figure 3

Results

For the first data set, the LC-MS data correlation analysis was done with the nanoESI data to get a r-value of 0.795, the LC-MS data with the femtoESI data to get a r-value of 0.923, and the nanoESI data with the femtoESI data to get a r-value of 0.781. The results of this method can be seen in table 1. The second method that was used on the first data set was the data analysis tool within excel (CORREL). This tool automatically calculated the correlation efficient, or r-value, for each comparison pair that was previously mentioned and yielded the same results as the previous method, as seen in table 2.

The same two methods of correlation analysis were applied to the second set of data provided. The results produced from the Pearson method were 0.770 for LC-MS with nanoESI, 0.961 for LC-MS with femtoESI, and 0.742 for nanoESI with femtoESI. These results can be seen in table 3. This first round of results was done up until there were no longer values provided for all three methods of analysis (LC-MS, nanoESI, femtoESI). After this point, there was no longer data provided for the femtoESI method, so the Pearson method was used again on the remaining values for only LC-MS and nanoESI. The produced value was 0.841 for LC-MS with nanoESI. This result can be seen in table 4.

The data analysis tool was also used on the second data set to further validate the results of the Pearson method. The results produced for the portion of the data set with all three methods included were 0.770 for LC-MS with nanoESI, 0.961 for LC-MS with femtoESI, and 0.742 for nanoESI with femtoESI. These results can be seen intable 5.

The second half of the second data set was also analyzed using this function in excel to validate the results for only LC-MS and nanoESI. The produced result was 0.841 for LC-MS with nanoESI. This result can be seen in table 6.

The polymer provided, $C_{31}H_{53}N_3O_{16}$, was created in *ChemDraw*, which was able to provide the exact mass (723.34$_{g/mol}$) as well as the molecular weight (723.77$_{g/mol}$) of the structure as seen in figure 2. The *MatLab* code was then written to cross-examine the value of the molecular weight given by *ChemDraw* (724.27$_{g/mol}$).

The mass spectrometry spectrum showed that there was in fact a peak at 724$_{m/z}$ when the spectrum was zoomed in on the point. This peak indicated that the previous programs, *MatLab* and *ChemDraw* were mostly accurate in their molecular mass calculations.

Conclusions

With the first data set provided, it seemed as though the highest correlation was present between the LC-MS data and the femtoESI data as it had the highest correlation coefficient value. On the other end, the nanoESI data and femtoESI data seemed to have the lowest correlation coefficient value which indicates a low correlation between the two sets of data.

With the second set of data provided, the LC-MS and femtoESI correlation was also the strongest with a value of 0.961. The lowest correlation for this data set seemed to be between nanoESI and femtoESI with a value of 0.742.

Both methods of calculating the correlation produced the same results, further strengthening the conclusion that the strongest correlation was between the LC-MS data and the femtoESI data with resulting values being only about 0.08 away from having perfect correlation. It should be noted, however, that the CORREL function within excel is basically the same as the Pearson method for calculating correlation, so there is room for error as the methods should be producing the same results and the Pearson function has been known to have some rounding errors in earlier versions of excel.

After comparing the molecular masses given by both *ChemDraw* and *MatLab*, it could be determined that the molecular mass was fairly accurate for the given polymer, $C_{31}H_{53}N_3O_{16}$, as the results from *ChemDraw* gave a molecular weight value of 723.77$_{g/mol}$ and *MatLab* produced a molecular weight of 724.27$_{g/mol}$. With the masses being so close to one another, it can be concluded that the accuracy is quite good.

Using the molecular masses produced by *MatLab* and *ChemDraw*, the mass spectrometry spectrum was able to further prove that the masses provided by the programs were quite accurate. The peak appearing at about 724$_{m/z}$ is especially close to the *MatLab* molecular mass of 724.27$_{g/mol}$ which seems to point towards *MatLab* being the more accurate of the two programs, even while the *ChemDraw* value was extremely close.

Acknowledgements

A huge thank you to my research advisor, Professor Anyin Li, for all of his help and for giving me the opportunity to be a part of his research group! Another thank you to my friends and family who have supported me through my studies!

References

1. Fonseca, M. 5 Things biomedical researchers need to know about correlation analysis. Editage Insights. https://www.editage.com/insights/5-things-biomedical-researchers-need-to-know-about-correlation-analysis (accessed 2024-10-16).
2. Janse, R. J.; Hoekstra, T.; Jager, K. J.; Zoccali, C.; Tripepi, G.; Dekker, F. W.; van Diepen, M. Conducting Correlation Analysis: Important Limitations and Pitfalls. Clinical Kidney Journal 2021, 14 (11), 2332–2337. https://doi.org/10.1093/ckj/sfab085.
3. Professor Anyin Li; Chemistry Department, University of New Hampshire https://www.unh.edu/, Durham, NH. Personal communication, 2025.