Experiment 4: Size exclusion chromatography

The primary objectives of this experiment were to separate four known proteins and one unknown protein using size exclusion chromatography as well as to determine the identity of the unknown from the data collected. In order to accomplish this we learned how to pack a chromatography column with size exclusion resin, run the column, measure the absorbances using UV-Vis spectrophotometry, and analyze the data using Mathematica.

Methods:

1. A 1 x 30 cm BioRad column was packed using 28.5 mL of Superdex 75 Prep Grade resin slurry to a height of 24.0 cm.
2. We prepared an elution buffer consisting of 150.0 mM NaCl and 50.0 mM sodium phosphate at a pH of 7.0.
3. The column was run at a volumetric flow rate of 0.33 mL/min after we added 80 µL of sample containing: Blue Dextran, Ovalbumin, Myoglobin, Vitamin B12, and an unknown.
4. Fraction collection was initiated at 14 minutes.
5. In total 48 fractions were collected with volumes of about 285 µL. Each fraction took approximately 75 seconds.
6. We used both an A and a C Shimadzu Biospec 1600 spectrophotometer to measure the spectra from 220 to 720 nm.

Calculations:

```
(*Data Analysis Calculations*)
(*Calculating the Capacity Factor (k'): k'=(tR-tM)/tM *)
k'Ovalbumin = (1815. s - 1365. s)/1365. s
```

Out[2]= 0.32967

```
(*Calculating Unknown Molecular Weight (MW)*)
MW = 10^((1.20879*-1.71798)+5.75873)
```

Out[3]= 4808.98

```
(*Calculating the Selectivity Factor (α): αAB=k'A/k'B *)
αOvalbumin/Myoglobin = 0.32967/0.76923
```

Out[4]= 0.428571

```
(*Calculating the Resolution (R): RAB=2ΔtR/wavg *)
RBlueDextran/Ovalbumin = (1815. s - 1365. s)/((375 s+525 s+525 s+525 s+600 s)/5)
```

Out[9]= 0.882353

In[11]:=
```
(*Calculating the Number of Theoretical Plates (N): N=16(tR/w)^2 *)
NBlueDextran = 16 * (1365./375.)^2
```

Out[11]= 211.994

In[12]:=
```
(*Calculating the Height Equivalent to the Theoretical Plate (H): H=L/N *)
HBlueDextran = (24 cm)/211.994
```

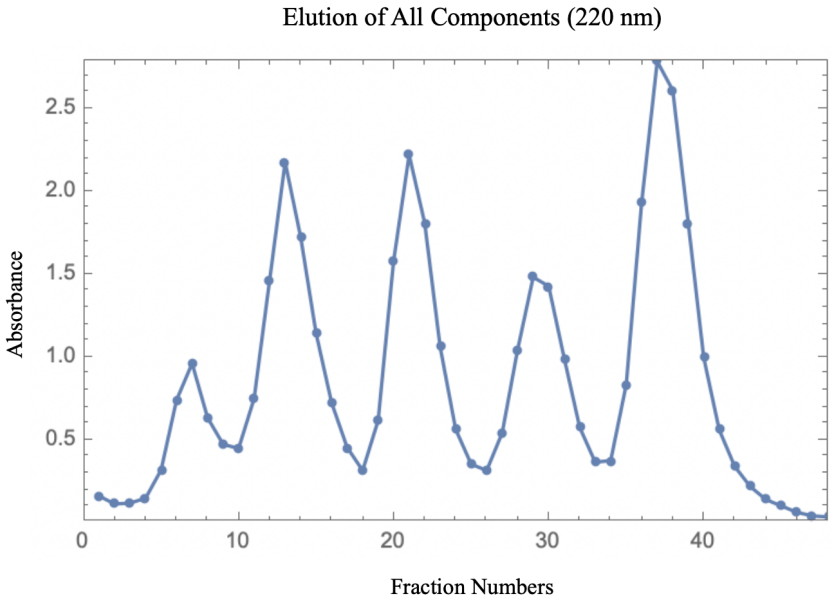
Out[12]= 0.113211 cm

Data Analysis:

Checking for missing data:

```
In[25]:= indx = Table[i, {i, 1, 48, 1}]
         absData = {};
         fileNumbers = {};

         For[i = 1, i ≤ Length[indx], i ++,
           If[indx[[i]] < 10,
            fname = "/Users/shaneabbley/MW6_grpB_" <> "0" <>
              ToString[indx[[i]]] <> ".txt",
            fname = "/Users/shaneabbley/MW6_grpB_" <> ToString[indx[[i]]] <>
               ".txt";];
           Print[fname];
           absVals = Import[fname, "Table"][[All, 2]];
           If[Length[absVals] == 251,
            absData = Append[absData, absVals];
            fileNumbers = Append[fileNumbers, indx[[i]]],
            Print[Style["Warning:Missing Data", Red]];]];

Out[25]= {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,
          19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
          34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48}
```

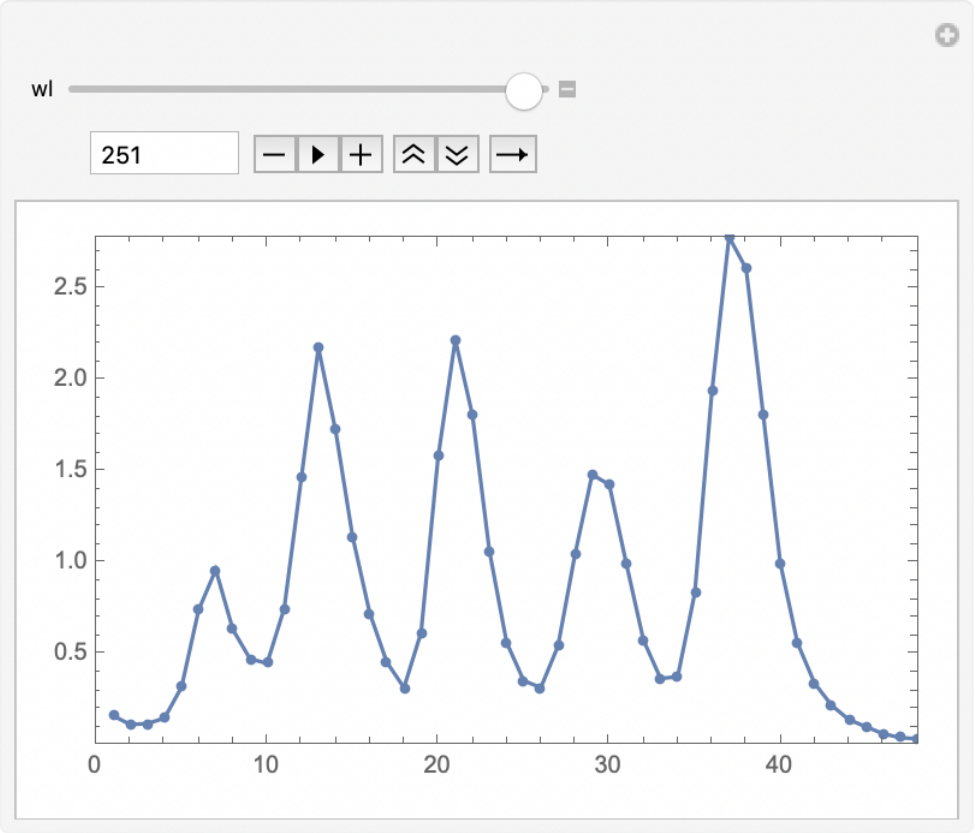
Ideal wavelength for visualizing elution of all components:



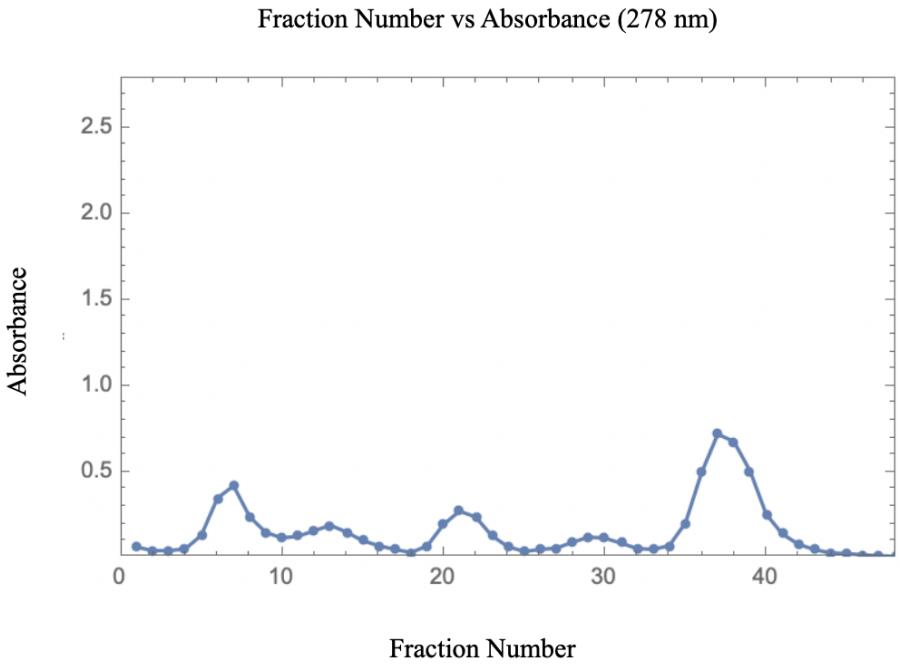
Used this manipulation of a line plot to determine noteworthy wavelengths, which show absorbances at various Fraction Numbers:

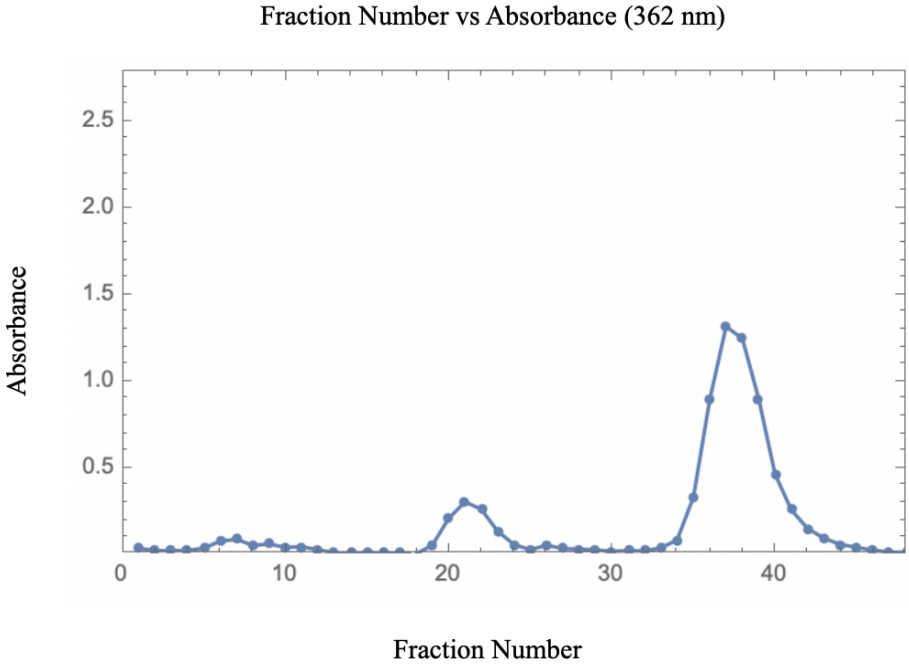
In[142]:=
```
Findwavelengths =
 Manipulate[ListPlot[Transpose[{indx, absData[[All, wl]]}],
   PlotRange → {{0, Max[indx]}, {0.01, Max[absData]}},
   Frame → True, Axes → False, Joined → True, Mesh → All],
  {wl, 1, 251, 1}]
```

Out[142]=

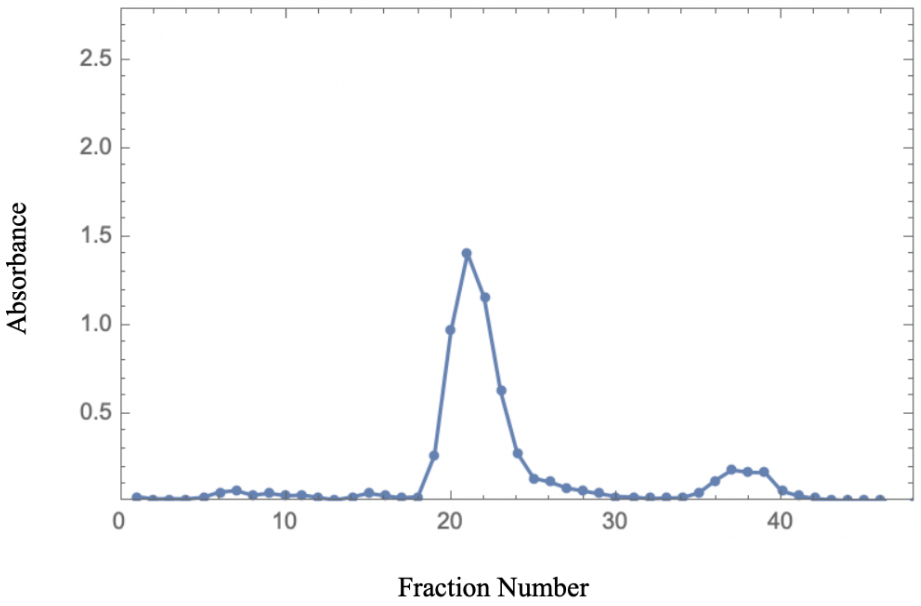


From this I found notable plots at the following wavelengths:

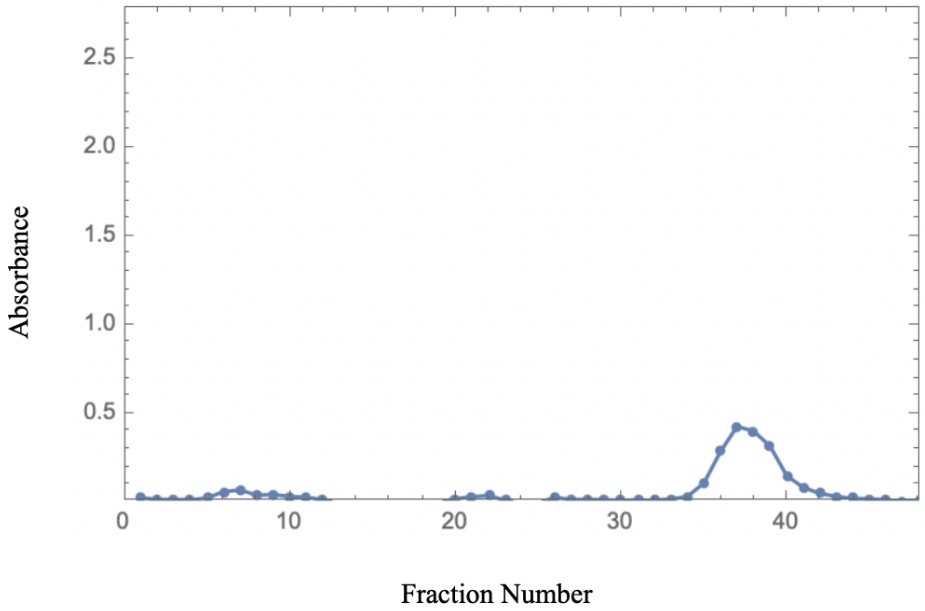




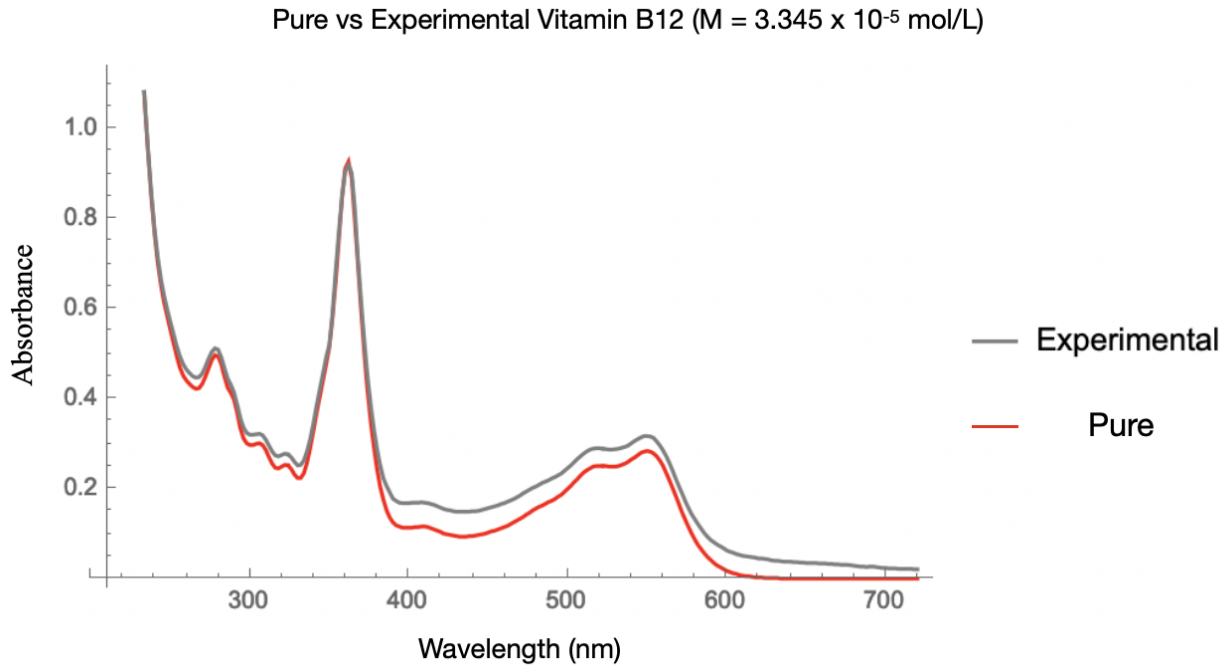
Fraction Number vs Absorbance (406 nm)



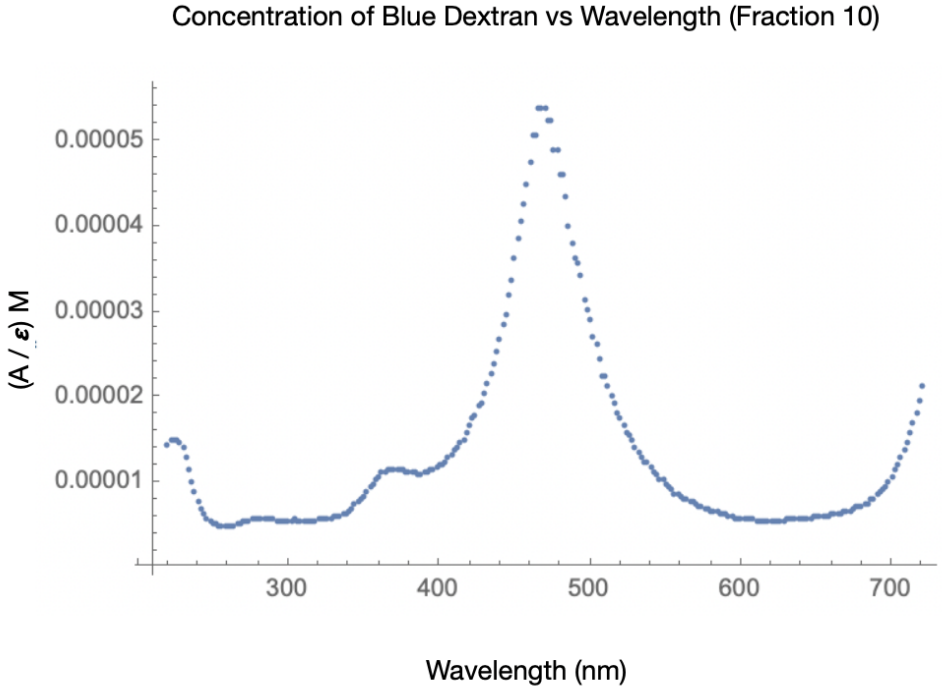
Fraction Number vs Absorbance (548 nm)



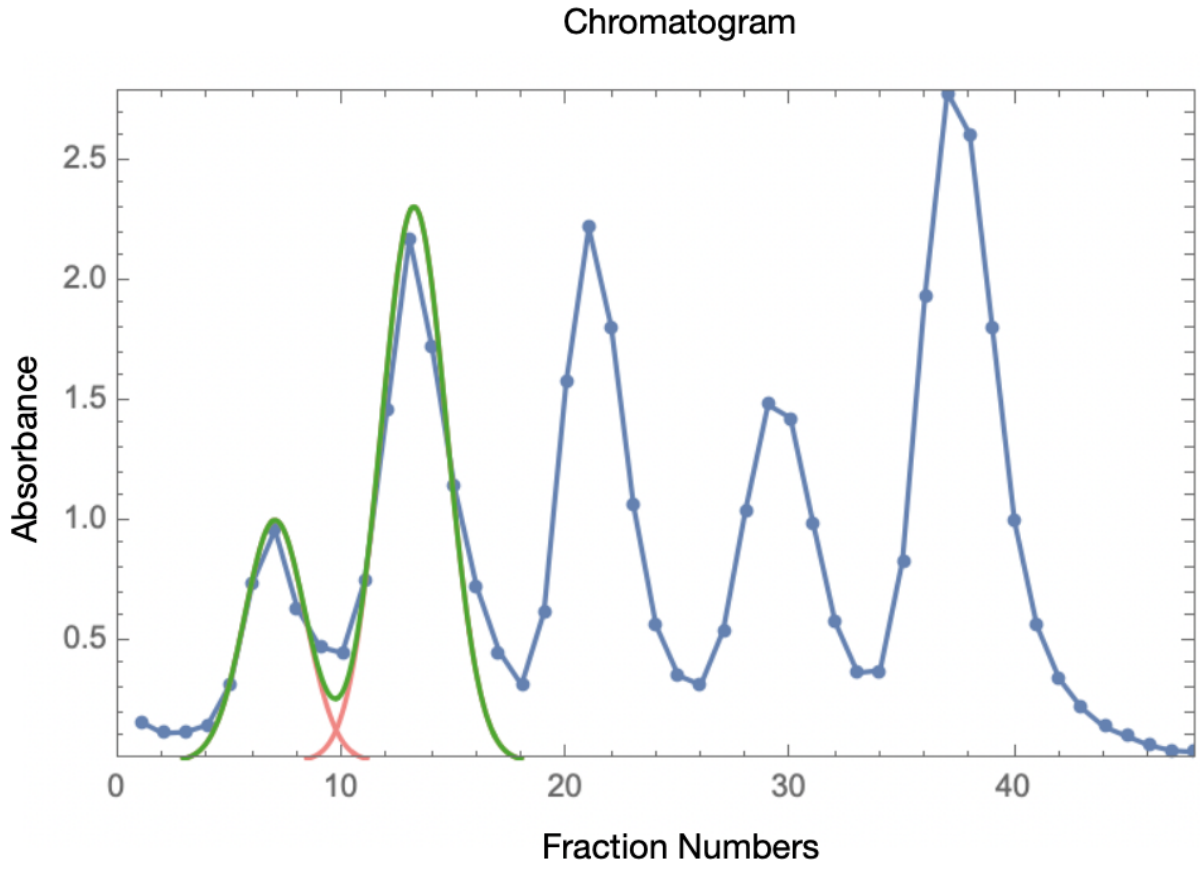
Fraction 39 appeared very pure so I overlaid it with the pure protein for Vitamin B12 and calculated the concentration from the absorption and corresponding molar extinction coefficient.



No fractions had appreciable absorbances for two solutes, but I chose fraction 10 to make a concentration vs wavelength plot.



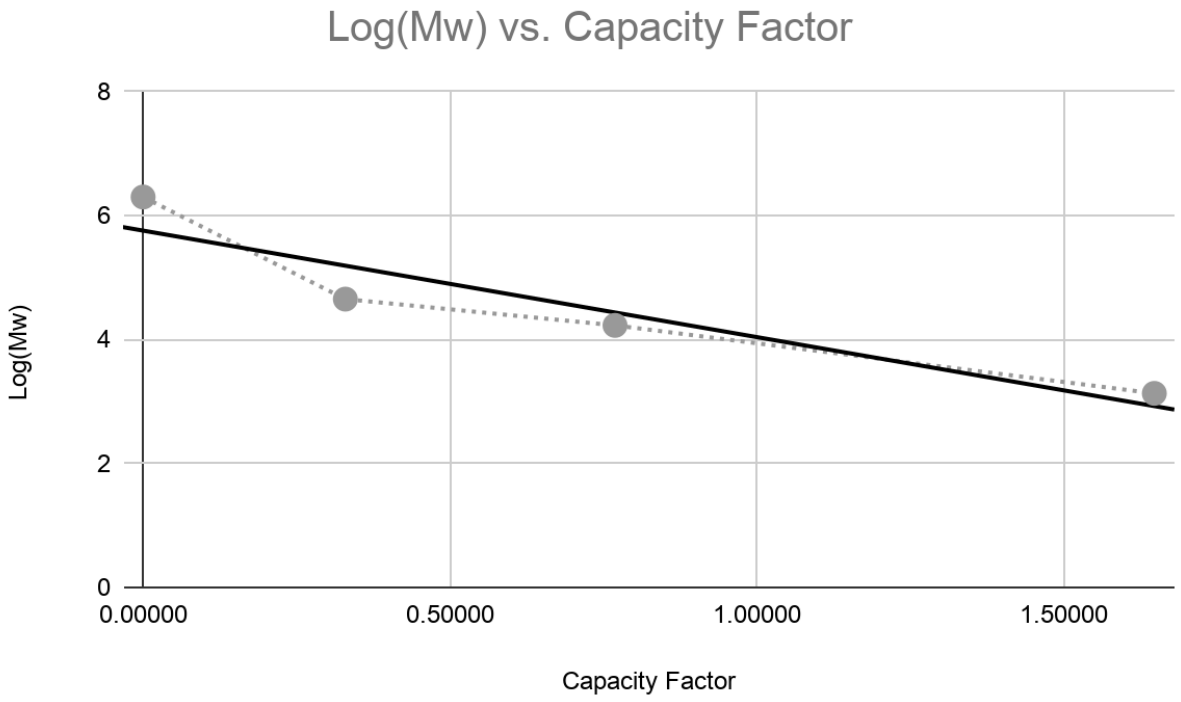
Very little overlap in our chromatogram. Even the two most overlapping curves, seen below, in green still are fairly well separated and no deconvolution is required for this qualitative analysis.



Retention Times calculated from the dead time of 1260 seconds added to the time taken for half of the solute to elute. The Peak width was determined using Gaussian functions, similar to those seen above.

	Retention Time (s)	Peak Width (s)	Capacity Factor
Blue Dextran (fractions 5-9)	1365	375	0.00000
Ovalbumin (fractions 10-16)	1815	525	0.32967
Myoglobin (fractions 19-25)	2415	525	0.76923
Unknown (27-33)	3015	525	1.20879
Vitamin B12 (35-42)	3615	600	1.64835

The linear best fit has an equation of $\hat{y} = -1.71798X + 5.75873$. The coefficient of determination is 0.871. The graph of the data presented above along with the best fit line looks as such:



Plugging the unknown's capacity factor into the best fit equation we get:

4808.98 g/mol

A matrix of selectivity factors:

	Ovalbumin	Myoglobin	Unknown	Vitamin B12
Ovalbumin	1.0	0.42857	0.27273	0.20000
Myoglobin	2.33333	1.0	0.63636	0.46667
Unknown	3.66667	1.57143	1.0	0.73333
Vitamin B12	5.00000	2.14286	1.36364	1.0

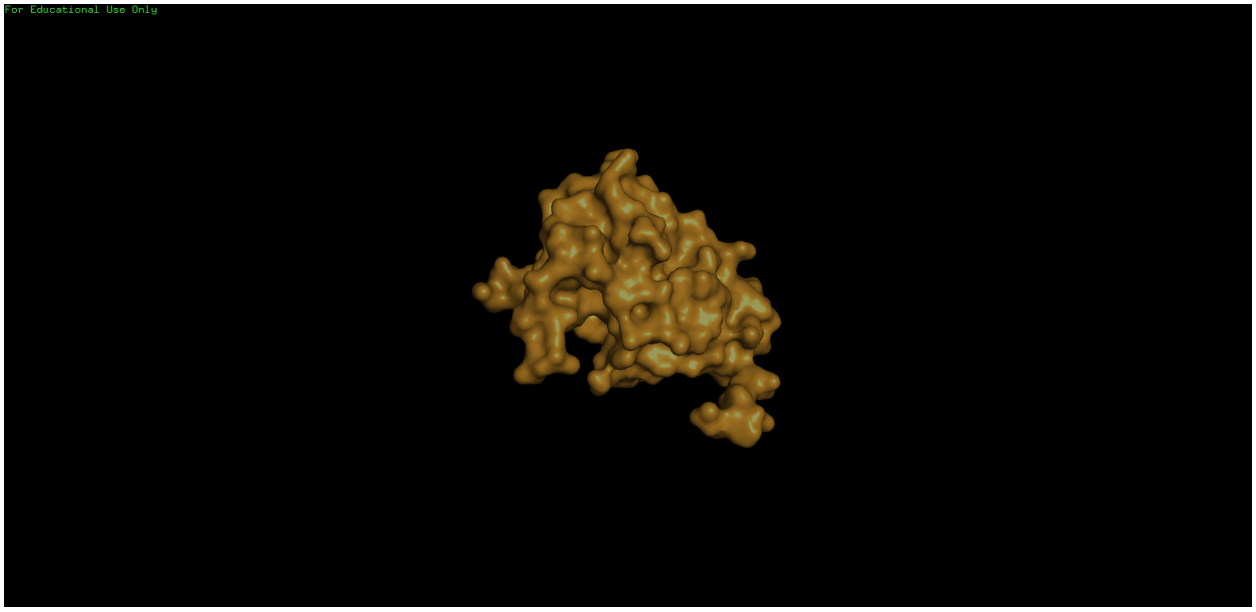
A matrix of resolutions:

	Blue Dextran	Ovalbumin	Myoglobin	Unknown	Vitamin B12
Blue Dextran	0	0.88235	2.05882	3.23529	4.41176
Ovalbumin	-0.88235	0	1.17647	2.35294	3.52941
Myoglobin	-2.05882	-1.17647	0	1.17647	2.35294
Unknown	-3.23529	-2.35294	-1.17647	0	1.17647
Vitamin B12	-4.41176	-3.52941	-2.35294	-1.17647	0

Data Analysis Part 19: A table of the theoretical plates and theoretical plate heights:

	Number of Theoretical Plates	Height Equivalent (cm)
Blue Dextran	211.99360	0.11321
Ovalbumin	191.22939	0.12550
Myoglobin	338.56000	0.07089
Unknown	527.68653	0.04548
Vitamin B12	580.81000	0.04132

Our unknown was most likely aprotinin:



Discussion:

It seems as though the quality of our collected data is sufficient for our purposes; however, there is much room for improvement in our experimental design as well as in our data analysis. The main factors that I analyzed to determine whether our objective was appropriately met were the quality of our chromatogram, the quantities of our resolutions, and the relative success of our data compared to other similar experiments.

First and foremost, our chromatogram presents to us the approximate fractions/elution times which our proteins eluted in. This shows the separation of our protein mixture after it has eluted from the column. It shows a clear separation of five unique bands. This is what we had expected and desired. The first peak eluted the most into fraction number seven and is indicative of the presence of Blue dextran 2000. Blue dextran was a key marker throughout this experiment and the data analysis. We knew going into the experiment that Blue dextran experienced little to no interaction with the pores of the Superdex 75 prep grade because it has a molecular weight of 2,000,000 g/mol, so we used it as an indication for when to start fractionation. In the data analysis Blue dextran was useful because we used its retention time as the void/solute elution time in calculations for the other proteins capacity factors. The second band overlapped with the Blue dextran band more than any other two bands, yet there is still clear separation between the two. The second band eluted mostly into fraction number thirteen. This fraction contained the duck egg protein ovalbumin, which has a molecular weight of 45,000 g/mol and thus entered some of the pores of the Superdex 75 prep grade resin. The third band was found in fraction 21 and appears to be the horse myoglobin protein, which has a molecular weight of 16,951 g/mol. This band was the best separated of them all and this can probably be attributed to the fact that it has a molecular weight that is quite far from that of ovalbumin and the unknown, which are the

closest proteins on each side. Fraction numbers 29 and 30 contained a majority of the unknown. The unknown's molecular weight was estimated to be 4,809 g/mol according to the best fit equation of our Log(M_W) vs capacity factor graph. This value is not particularly close to the actual value of 6,518 g/mol, but given our list of possible unknowns, it is the only reasonable choice. Furthermore, after analyzing the structure of aprotinin, we can see that the structure has various arm-like secondary structures that stick out of the protein. This would increase drag and interactions with other molecules and thus make the capacity factor slightly higher than expected for the given molecular weight. Further analysis of the partial specific volume shows that aprotinin is comparable to cytochrome c, which has a M_W of 12,398. This means that the effective volume of aprotinin is greater than predicted due to its interactions with the solute, the pores, or other proteins.

Several other factors also lead to the error observed in the estimated molecular weight of aprotinin. More proteins are required to be separated to achieve better regression analysis. This is because the sample size is simply too small and there are not enough degrees of freedom for this experiment. Another limitation of our data comes from the size of the column. Better resolution would be attained using a similar set up, but increasing the size of the column to approximately 50 cm, as we specified in our prelab. This would allow for greater separation and thus peaks would be farther apart, but they would also be broader. This could be a problem because the peaks were already slightly broader than ideal and this error is more or less due to the longitudinal diffusion. We could also have decreased the volumetric flow rate at the cost of time. As seen in the table of resolutions none of these proteins were sufficiently separated for quantitative analysis. Fortunately, our analysis was fairly straightforward and forgiving, such that

we could save time by running our procedure as we did and still get data that points us in the direction of the correct unknown.

As compared to other groups the two Superdex 75 prep grade columns proceeded with the fewest hitches. The two Superdex columns yielded better results than the Sephacryl 5200-HR and the Toyopearl HW55 superfine. The Toyopearl has comparable resolution, but is more challenging to work with and the Sephacryl 200-HR had minor difficulty with the smaller residues. Also the column heights chosen (30 cm and 50 cm) were both sufficient for this experiment. Although the longer 50 cm column beds did take significantly more time for comparable resolution than our own 24 cm bed in a 30 cm column. The volumetric flow rates were specifically selected for the chosen columns, so it is difficult to compare them directly; however, small adjustments would alter the resolution and the peak width. Ideally, there is an optimization where the minimal weighted effects of each are obtained and thus the most efficient combination of resolution and time. Overall, the greatest factor in the success of the experiment appeared to be the resin.

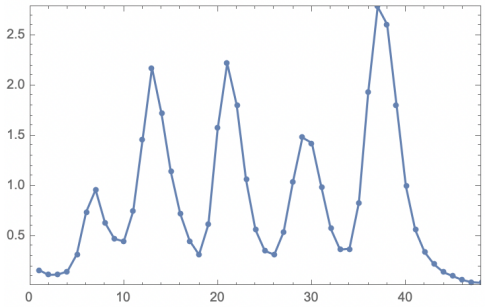
Using Mathematica proved invaluable for the analysis of some of the data in this lab. Plotting the absorbance and molar extinction coefficient data from the large quantities of collected data files and pure protein sample data files saved a lot of time. The ListPlot function clearly organized the imported data into easily understandable chromatogram plots. The matrix functions in mathematica allowed for quick calculations of the selectivity factors and resolution between the corresponding peaks. These calculations were simple enough such that mathematica would save no more than a few minutes, but in theory, as the matrix size (number of proteins) increases the amount of time saved would increase exponentially. The manipulate function applied to the plots made it easier to determine the identity of the samples by quickly adjusting a

wavelength slider that changes the absorbance as a function of fraction number. Lastly the simple gaussian curves plotted over the chromatogram allowed for a greater visualization of the overlap of the peaks and a good line such that the peak widths could be easily determined.

In the end, the Superdex 75 prep grade sufficiently separated the peaks of our protein and provided resolution that adequately led to the conclusion that our protein sample contained the unknown protein aprotinin. Using mathematica and the data acquired from other methods we determined that our procedure was well designed with few key errors.

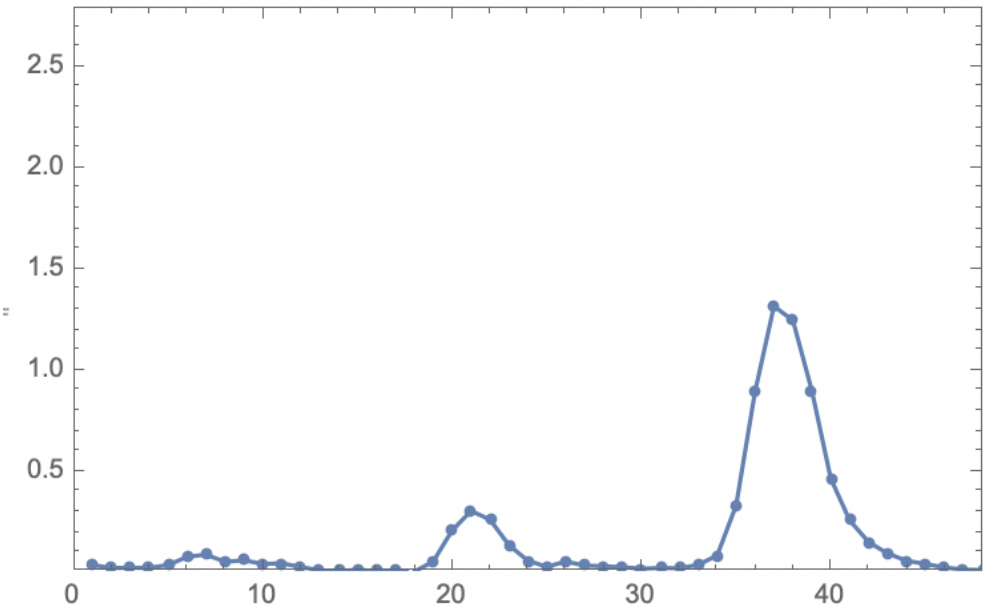
Appendix:

```
$Path;

Directory[]

FileNames["*.txt"];

indx = Table[i, {i, 1, 48, 1}]
absData = {};
fileNumbers = {};

For[i = 1, i ≤ Length[indx], i ++,
  If[indx[[i]] < 10,
   fname = "/Users/shaneabbley/MW6_grpB_" <> "0" <> ToString[indx[[i]]] <> ".txt",
   fname = "/Users/shaneabbley/MW6_grpB_" <> ToString[indx[[i]]] <> ".txt";];
  Print[fname];
  absVals = Import[fname, "Table"][[All, 2]];
  If[Length[absVals] == 251,
   absData = Append[absData, absVals];
   fileNumbers = Append[fileNumbers, indx[[i]]],
   Print[Style["Warning:Missing Data", Red]];]];

(*5 Elution of All Components*)
plot220 = ListPlot[Transpose[{indx, absData[[All, -1]]}],
  PlotRange → {{0, Max[indx]}, {0.01, Max[absData]}},
  Frame → True, Axes → False, Joined → True, Mesh → All]
```

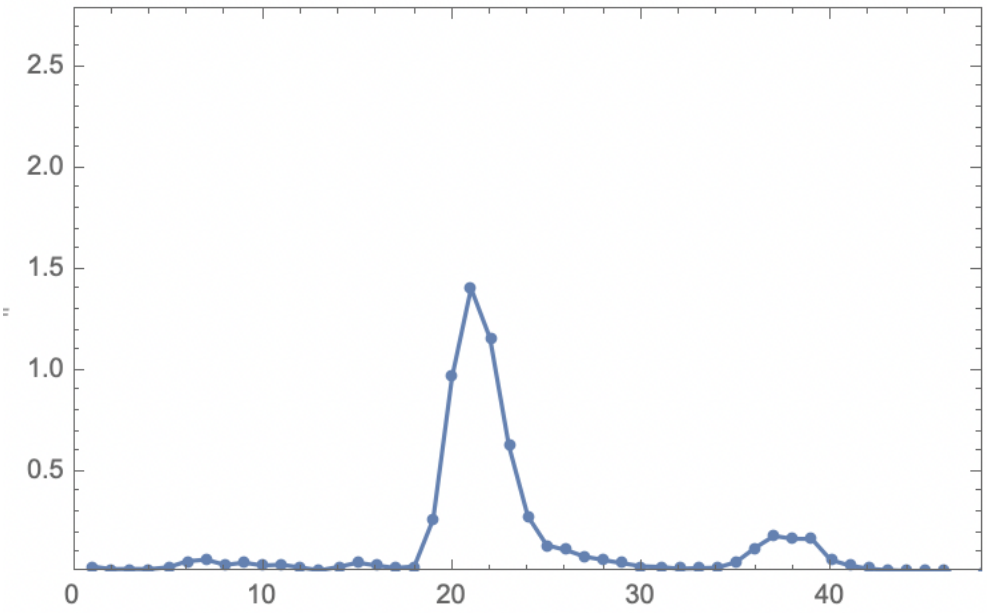



```
Manipulate[
 Show[plot220, Plot[{c1 * PDF[NormalDistribution[m1, s1], x], c2 * PDF[NormalDistribution[m2, s2], x],
    c1 * PDF[NormalDistribution[m1, s1], x] + c2 * PDF[NormalDistribution[m2, s2], x]}, {x, 0, 48},
   PlotStyle → {Pink, Pink, Darker[Green]}, PlotRange → {{0, 48}, {0.01, Max[absData]}}]], {{c1, 5.0}, 1, 10, 0.1},
 {{c2, 5.0}, 1, 10, 0.1}, {{m1, 5.6}, 0, 30, 0.1}, {{m2, 7.6}, 0, 30, 0.1}, {{s1, 1.3}, 1, 3, 0.05}, {{s2, 1.4}, 1, 3, 0.05}]

Findwavelengths = Manipulate[ListPlot[Transpose[{indx, absData[[All, wl]]}],
   PlotRange → {{0, Max[indx]}, {0.01, Max[absData]}}, Frame → True, Axes → False, Joined → True, Mesh → All], {wl, 1, 251, 1}]
```

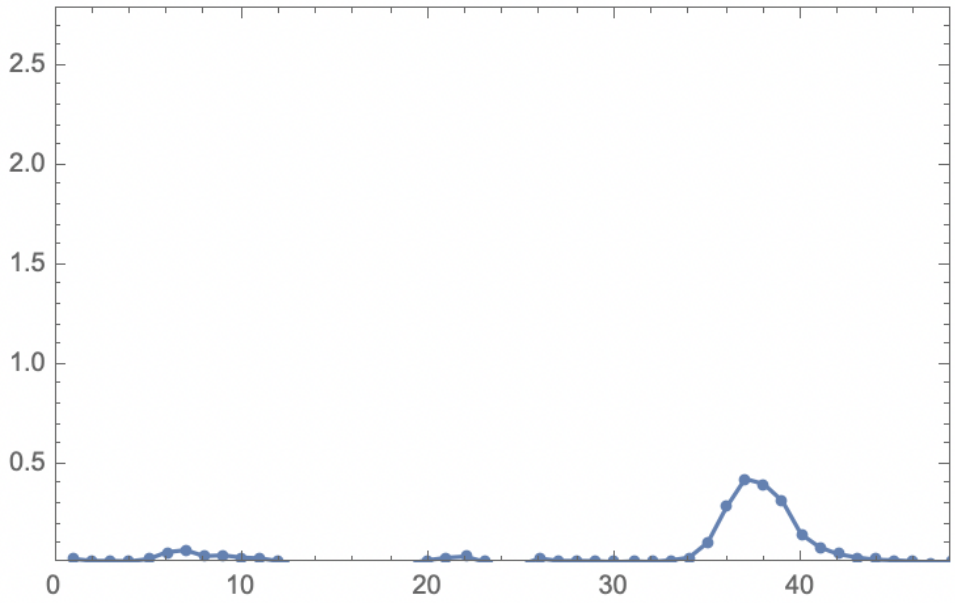
```
plot179 = ListPlot[Transpose[{indx, absData[[All, 179]]}],
  PlotRange → {{0, Max[indx]}, {0.01, Max[absData]}},
  Frame → True, Axes → False, Joined → True, Mesh → All]
```



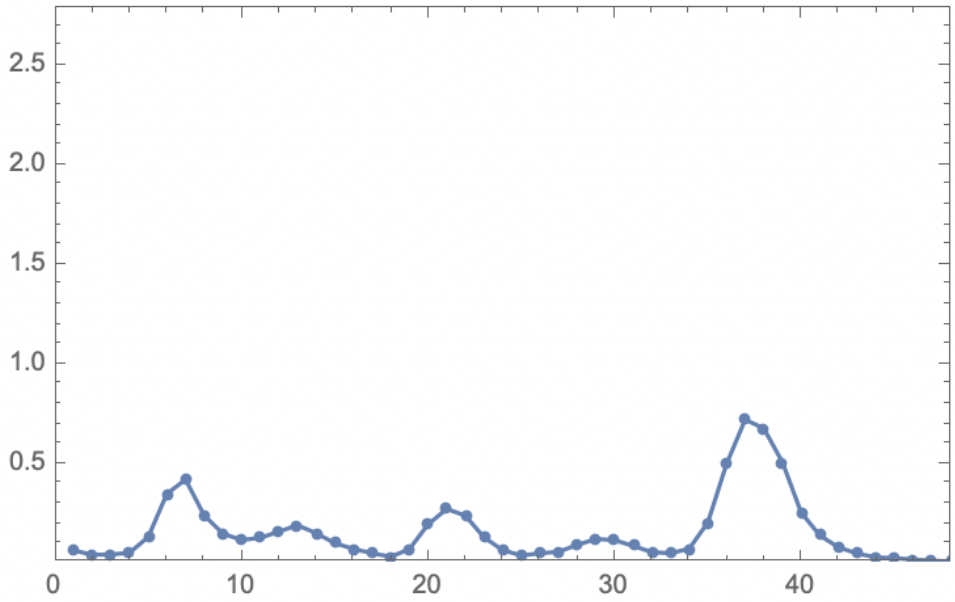

```
plot157 = ListPlot[Transpose[{indx, absData[[All, 157]]}],
  PlotRange → {{0, Max[indx]}, {0.01, Max[absData]}},
  Frame → True, Axes → False, Joined → True, Mesh → All]
```



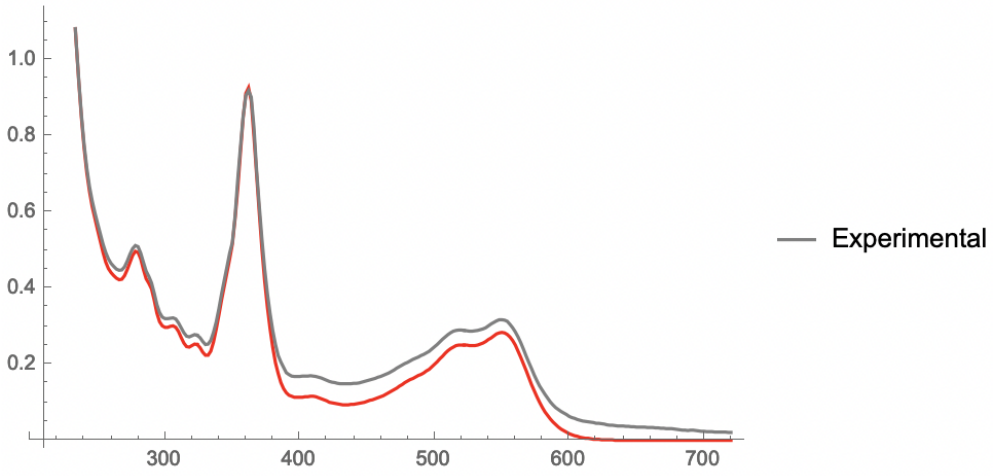
```
plot86 = ListPlot[Transpose[{indx, absData[[All, 86]]}],
  PlotRange → {{0, Max[indx]}, {0.01, Max[absData]}},
  Frame → True, Axes → False, Joined → True, Mesh → All]
```



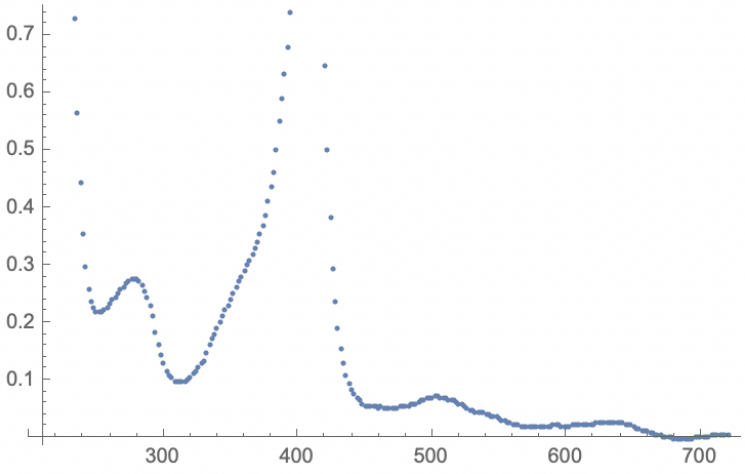

```
plot221 = ListPlot[Transpose[{indx, absData[[All, 221]]}],
  PlotRange → {{0, Max[indx]}, {0.01, Max[absData]}},
  Frame → True, Axes → False, Joined → True, Mesh → All]
```

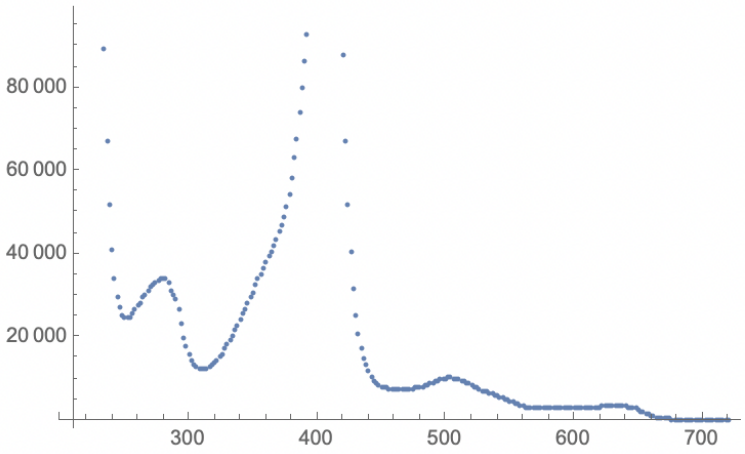


```
data = Import["/Users/shaneabbley/MW6_grpB_39.txt", {"Data", {All}, {1, 2}}];
datagraph = ListPlot[data, Joined -> True, PlotStyle -> Gray,
    PlotLegends → {"Experimental"}];
pure = Import["/Users/shaneabbley/Spectra/220-720/vitaminB12_eps_251.txt",
    {"Data", {All}, {1, 2}}];
w = pure[[All, 1]];
e = pure[[All, 2]];
z = e / 29 650;
final = Transpose[{w, z}];
puregraph = ListPlot[final, Joined -> True, PlotStyle → Red];
Comparison = Overlay[{puregraph, datagraph}]
```

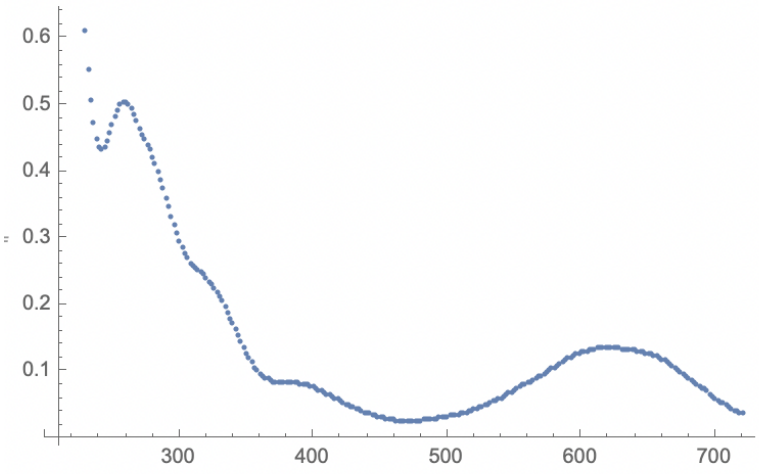


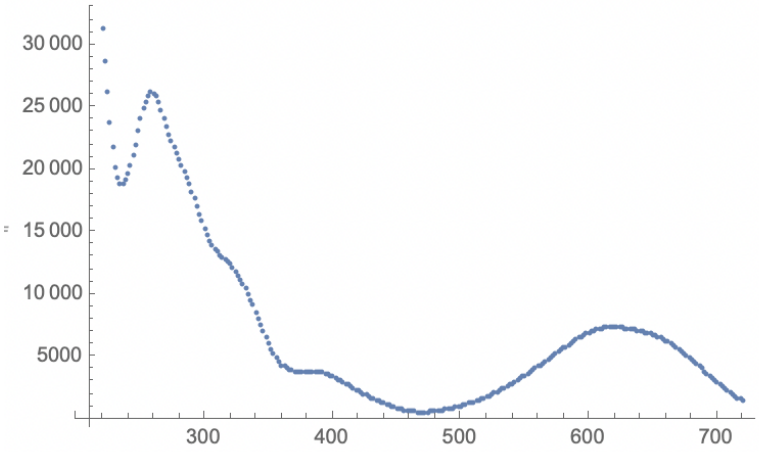
```
data = Import["/Users/shaneabbley/MW6_grpB_21.txt", {"Data", {All}, {1, 2}}];
pure = Import["/Users/shaneabbley/Spectra/220-720/myoglobin_eps_251.txt",
   {"Data", {All}, {1, 2}}];

ListPlot[data]
ListPlot[pure]
```





```
data = Import["/Users/shaneabbley/MW6_grpB_07.txt", {"Data", {All}, {1, 2}}];
ListPlot[data]
pure = Import["/Users/shaneabbley/Spectra/220-720/bluedext_eps_251.txt",
   {"Data", {All}, {1, 2}}];
ListPlot[pure]
```





```
data = Import["/Users/shaneabbley/MW6_grpB_13.txt", {"Data", {All}, {1, 2}}];
pure = Import["/Users/shaneabbley/Spectra/220-720/ovalbumin_eps_251.txt",
    {"Data", {All}, {1, 2}}];
w = pure[[All, 1]];
new = pure[[All, 2]];
done = new * (2.59 * 10^-5);
ok = Transpose[{w, done}];
ListPlot[data, PlotRange → {{220, 720}, {0, 0.70}}]
ListPlot[ok, PlotRange → {{220, 720}, {0, 0.80}}]
```