ClpX is a bacterial protease that uses unfoldase mechanisms to degrade substrates when it is complexed with ClpP, the proteolytic portion of the protease complex. ClpP forms a barrel with a central channel the substrates are fed through. ClpX sits on top of that barrel and uses the N-terminal side of its hexameric structure to recognize and bind to substrates. From there, the substrates get pulled on through an ATP-driven power stroke that unfolds the substrates so they can more easily be degraded by ClpP.

This script uses data generated by ClpX variants, each with a different single site mutation on the N-terminus, to observe and analyze the data for any distinct trends or effects on the variants ability to bind and degrade substrate.

The script first looks at data generated from a fluorescence based binding assay. In this assay, ClpX and ClpP are incubated together with a peptide of the first 76 residues of the substrate MqsA, and then spun with centrifugation in 100 kDa molecular weight cut-off columns. It is predicted that ClpX recognizes MqsA at its own N-terminus so those first 76 residues contain this region. The substrate also has a GFP fused to it so that it can be measured using a fluorometer. If ClpX or variants can bind to GFP-MqsA 1-76, the substrate will remain bound after the spin and the amount of fluorescence remaining after the spin is a direct measure of how well ClpX or variants bound the substrate. From the analysis, the ClpX A30S mutation had the highest retention rate of the GFP-MqsA 1-76 substrate. It appears that replacing the 30th position with a polar residue allows the unfoldase to more tightly bind the substrate. This specific mutant only had two replicates and the others had 5, so this could lead to the high error bar for the mutant. However, even at the low end of the error bar, it is still higher than WT. All the other mutations seem to have no effect at all compared to WT which is interesting because they are all closer to the beginning of the ClpX N-terminus than A30S.

The next dataset the script analyzes is the ATP hydrolysis data set. This is an important data set to observe because ClpX is an ATP-driven unfoldase so in order to confirm that the mutants are functional, they have to be able to hydrolyze ATP. ClpX is incubated with ATP for 15 minutes and a sample is taken every 5 minutes to track the amount of free phosphate, using malachite green phosphate detection buffer. The amount of free phosphate measured is a measurement of the amount of ATP being hydrolyzed. Although the binding of substrate has no correlation to ATP hydrolysis, the degradation of substrates is directly linked to ATP hydrolysis as it is what drives the unfolding of the substrate for ClpP degradation. Based on the plot generated using the script, it is apparent that the L13A mutation has the highest ATP hydrolysis rate and H23A has a modest increase. It is also apparent that L12S has the lowest. After seeing these results, it is expected that the higher the ATP hydrolysis, the more degradation of substrate you would expect. It is worth noting here that the A30S mutation has no data on this graph, that is because it has not yet been tested in this assay.

The last dataset that was analyzed was the MqsA-NTD degradation data set. This data was generated by incubating ClpX and ClpP together with MqsA-NTD which is the first 76 amino acids of MqsA without any additional fused tags. The incubation ran for 3 hours and a gel was run. Using densitometry, the gel bands were quantified and the resulting amount of substrate remaining from time 0 to time 3 hours was calculated. The plot generated from the script showed that H23A had the highest amount of degradation after the 3 hours. The other ClpX variants tested did not have much of an effect on degradation compared to WT ClpX. The L13A and A30S mutations were not tested in this assay yet which is unfortunate for this analysis because the plot would be incredibly informative to match the high binding of A30S or the high ATP hydrolysis of L13A with the actual degradation rate of the substrate.

The datasets were then combined in each possible combination of pairs in order to further check for any correlations. The pairs were, Binding/ATP, Binding/Degradation, and Degradation/ATP, and the datasets were joined by the common denominator of the mutant names.

First was the Binding/ATP dataset, which may not show any correlation as I previously mentioned because ATP hydrolysis doesn't affect substrate binding, only substrate unfolding, but can provide useful information on mutants that may be deficient or improved in both categories. Based on the plot generated from the script, it does not appear that there is any correlation between the mutants when comparing the results from the fluorescence based binding assay and the ATP hydrolysis assay.

Next was the Binding/Degradation dataset which should have a correlation because binding needs to take place in order for degradation to take place. This generated plot was interesting because it appeared that WT ClpX had the best combined binding and degradation results. It is also interesting to see the stark difference between the degradation of WT and H23A in this new plot. The degradation plot shows it well but when tying it together with the binding data, you can more clearly see that H23A is only binding the substrate slightly less than WT, yet is degrading it nearly 20% better. Something to keep in mind here is that some of the mutations can cause increased or decreased binding but that can actually have an effect on how well the substrate gets transferred from a bound position to an unfolded and transferred position for degradation. For instance, if one mutation binds the substrate too tightly, then it will have trouble releasing the substrate farther into the pore-channel to be unfolded and therefore will affect how well the substrate can get degraded, if at all.

The last combined dataset was the Degradation/ATP combo which should show the most direct correlation between the results of the two assays. ClpX can only unfold substrates

using the power of ATP hydrolysis and without that, the substrate will never be unfolded and could never be transferred to ClpP for degradation. From the plot of the combined dataset, it is clear that although H23A had a modest binding assay result, it is rapidly hydrolyzing ATP and transferring that hydrolysis into substrate degradation at a high rate. It is also interesting that the L12S mutation is hydrolyzing ATP more slowly than WT ClpX, yet having a slightly better substrate degradation rate. These two variants are so close together on the plot that the slight change may be within error for the two samples. Again, it is worth noting that some of the observed results could be from other factors taking place such as one mutation causing ATP hydrolysis to be upregulated even if the N-terminus residual change results in a worse substrate recognition.

There is still a lot of work left to be done on this project (3 years to be exact for myself), but the results that were obtained using this script could prove to be vital in directing the efforts going forward. This script has allowed me to look at the data in ways I have not yet been able to do so. By combining the datasets and then plotting, I was able to observe trends between assays that were only previously made by visually looking at raw numbers. All datasets had previously been plotted individually but the combination plots have opened up the door for truly mapping how each mutation is affecting each part of substrate recognition, binding, and degradation. Many more mutations still have to be made and will be added to these plots as I continue to progress on this work. I wish for the purpose of these current plots that I had been able to finish characterizing the A30S and L13A mutations in these assays because they both had individually promising results in the binding assay and hydrolysis assays respectively. Once that data is generated, these plots will play a vital role in observing the larger picture of ClpX N-terminus substrate recognition, binding, and degradation.