BIO_539_Final_Project_Piraino

Load in all the needed packages in order to run the upcoming scripts.

```
#install.packages("ggplot2")
library(ggplot2)
#install.packages("tidyverse")
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------------- t
```

```
## v tibble  3.0.1     v dplyr   0.8.5
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------------------------------------------------- tidyvers
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#install.packages("latexpdf")
library(latexpdf)
Binding_Data_For_R_Project_4 <- read_csv("Big_Data_Analysis/BIO539_Final_Project/Binding_Data_For_R_Pro
```

```
## Parsed with column specification:
## cols(
##   ClpX_Mutant = col_character(),
##   MqsA_1_76_GFP_Retained = col_double(),
##   SD = col_double()
## )
```

```
ATP_Hydrolysis_Data_for_R_Project_3 <- read_csv("Big_Data_Analysis/BIO539_Final_Project/ATP Hydrolysis
```

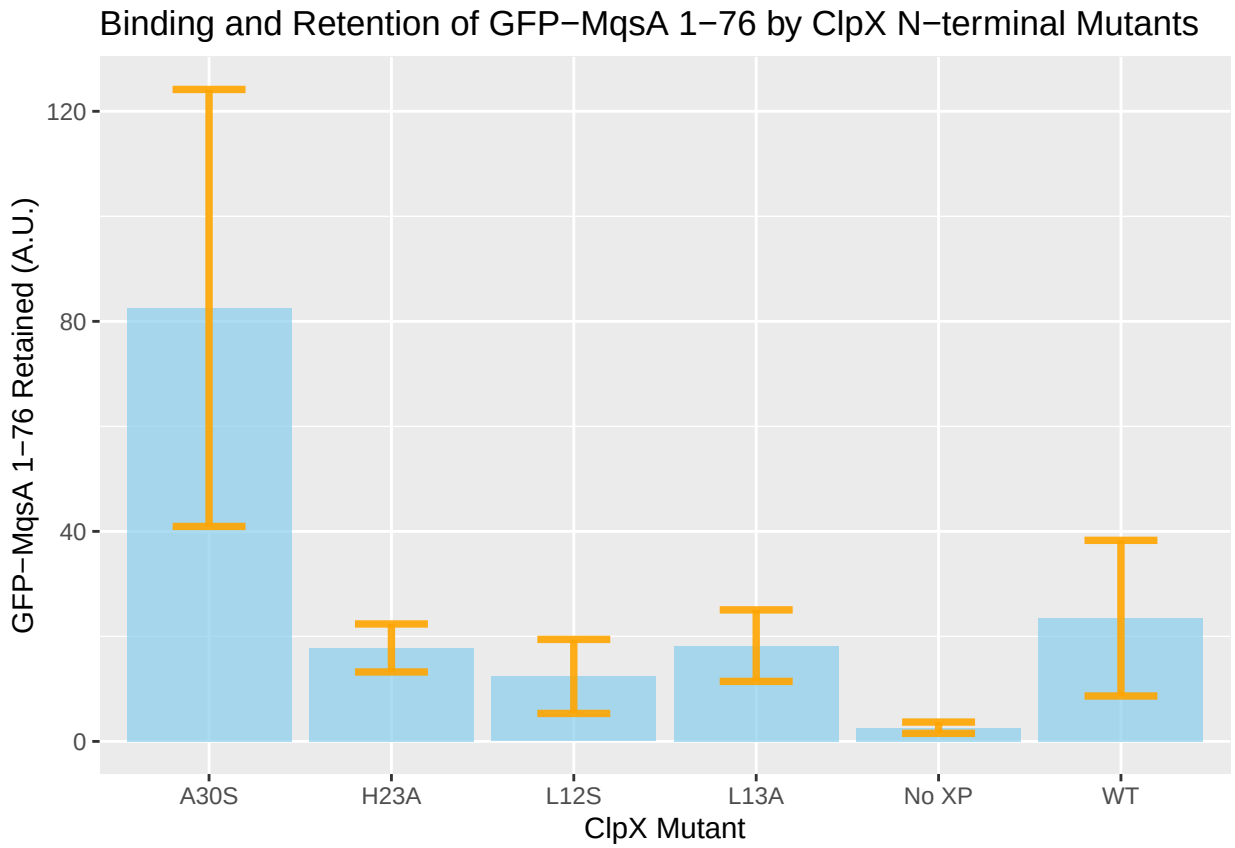
```
## Parsed with column specification:
## cols(
##   ClpX_Mutant = col_character(),
##   Avg_Hydrolysis = col_double(),
##   SD = col_double()
## )
```

```
MqsA_NTD_Degradation_R_Project_2 <- read_csv("Big_Data_Analysis/BIO539_Final_Project/MqsA_NTD_Degradati
```

```
## Parsed with column specification:
## cols(
##   ClpX_Mutant = col_character(),
##   MqsA_NTD_Degradation = col_double(),
##   SD = col_double()
## )
```

Plot the first data set, in this case, the fluorescence binding assay data

```
ggplot(Binding_Data_For_R_Project_4) +
  geom_bar( aes(ClpX_Mutant,MqsA_1_76_GFP_Retained), stat="identity", fill="skyblue", alpha=0.7) +
  geom_errorbar( aes(x=ClpX_Mutant, ymin=MqsA_1_76_GFP_Retained-SD, ymax=MqsA_1_76_GFP_Retained+SD),
                 width=0.4, color="orange", alpha=0.9, size=1.3) +
  labs(title = "Binding and Retention of GFP-MqsA 1-76 by ClpX N-terminal Mutants" ,
       x = "ClpX Mutant" , y = "GFP-MqsA 1-76 Retained (A.U.)")
```



#This plot shows the various ClpX mutants on the x-axis and the amount of fluorescence on the y-axis which is a direct representation of how much substrate was retained by the specific mutant protein. #This is the standard format for the remaining data sets as well. Each will be plotted with the ClpX mutants on the x-axis and the specific assay units on the y-axis. The other bit of code within the plot script determines the colors of the bars because this plot is a bar graph, and the width of the bars and intensity of the color. The plot also contains error bars which are determined by the y-min and y-max, which is the specific y-value for the mutant plus and minus the standard deviation for that mutant to determine the height above and below the y-value for the error bar to represent.

#You need to input the file name and the x- and y-axis names in the specified spots in order for this plot to work correctly.

#From the data you can see that the A30S mutation has the highest retention rate. Although this mutant has a high error bar, there are only two replicates run for this mutant while the others have 5 replicates and have dropped the high and the low. Even with the high error bar, the low end of the A30S error bar is still higher than all the other samples, showing that a polar mutation at the 30th residue plays an important role in the binding of GFP-MqsA 1-76. The other mutants all float moderately close to WT with no conclusions to be made.

Generating a summary of the fluorescence retained for the first assay

```
Binding_Data_For_R_Project_4 %>%
  summary(MqsA_1_76_GFP_Retained)
```

```
##  ClpX_Mutant        MqsA_1_76_GFP_Retained       SD
##  Length:6           Min.   : 2.612         Min.   : 1.07
##  Class :character   1st Qu.:13.733         1st Qu.: 5.13
##  Mode  :character   Median :18.017         Median : 6.93
##                     Mean   :26.175         Mean   :12.65
##                     3rd Qu.:22.163         3rd Qu.:12.88
##                     Max.   :82.551         Max.   :41.61
```

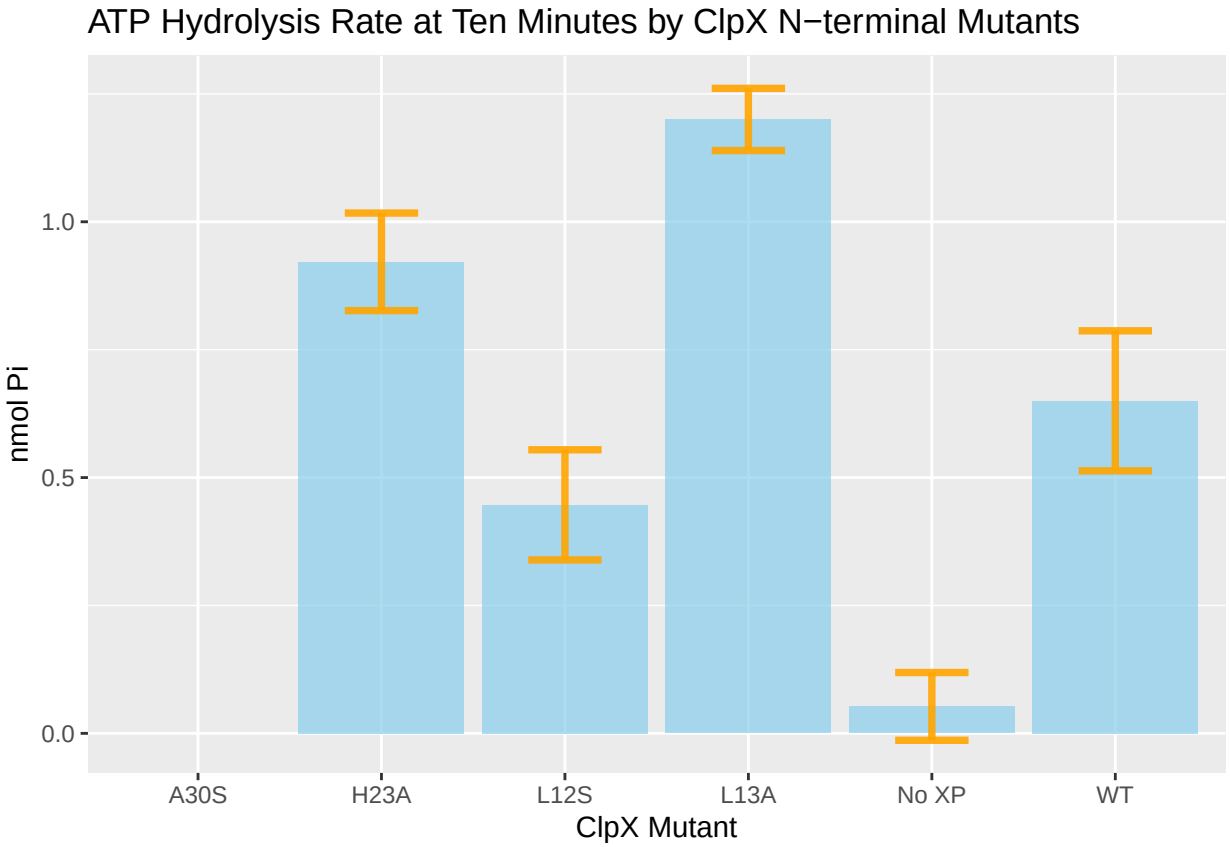
#This is not a necessary step but can be useful to visualize the mean fluorescence units retained across all mutants, the more data that is amassed.

#This process of plotting the data set and then calculating the summary will be repeated for the other data sets prior to combining data sets.

Plot the ATP hydrolysis data for the ClpX mutants

```
ggplot(ATP_Hydrolysis_Data_for_R_Project_3) +
  geom_bar( aes(ClpX_Mutant,Avg_Hydrolysis), stat="identity", fill="skyblue", alpha=0.7) +
  geom_errorbar( aes(x=ClpX_Mutant, ymin=Avg_Hydrolysis-SD, ymax=Avg_Hydrolysis+SD),
                 width=0.4, color="orange", alpha=0.9, size=1.3) +
  labs(title = "ATP Hydrolysis Rate at Ten Minutes by ClpX N-terminal Mutants" ,
       x = "ClpX Mutant" , y = "nmol Pi")
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```



#This plots the ClpX mutants on the x-axis and the average ATP hydrolysis in nmol Pi on the y-axis. Refer to the plotting of the fluorescence data to see what the other pieces of code here represent.

#The results of this plot show that the L13A mutant has a much higher ATP hydrolysis rate over the 10 minute hydrolysis reaction than the other mutations. H23A is also increased compared to WT Clpx. Here, A30S has no rate because it has not yet been tested for ATP hydrolysis. The data will be updated when that assay has been run. The hydrolysis of ATP is vital to the role of ClpX in unfolding substrates for degradation by ClpP but has not been shown to have any effect on actually recognizing and binding substrates which we will look at further on.

Calculate the summary of the average hydrolysis for all ClpX mutants

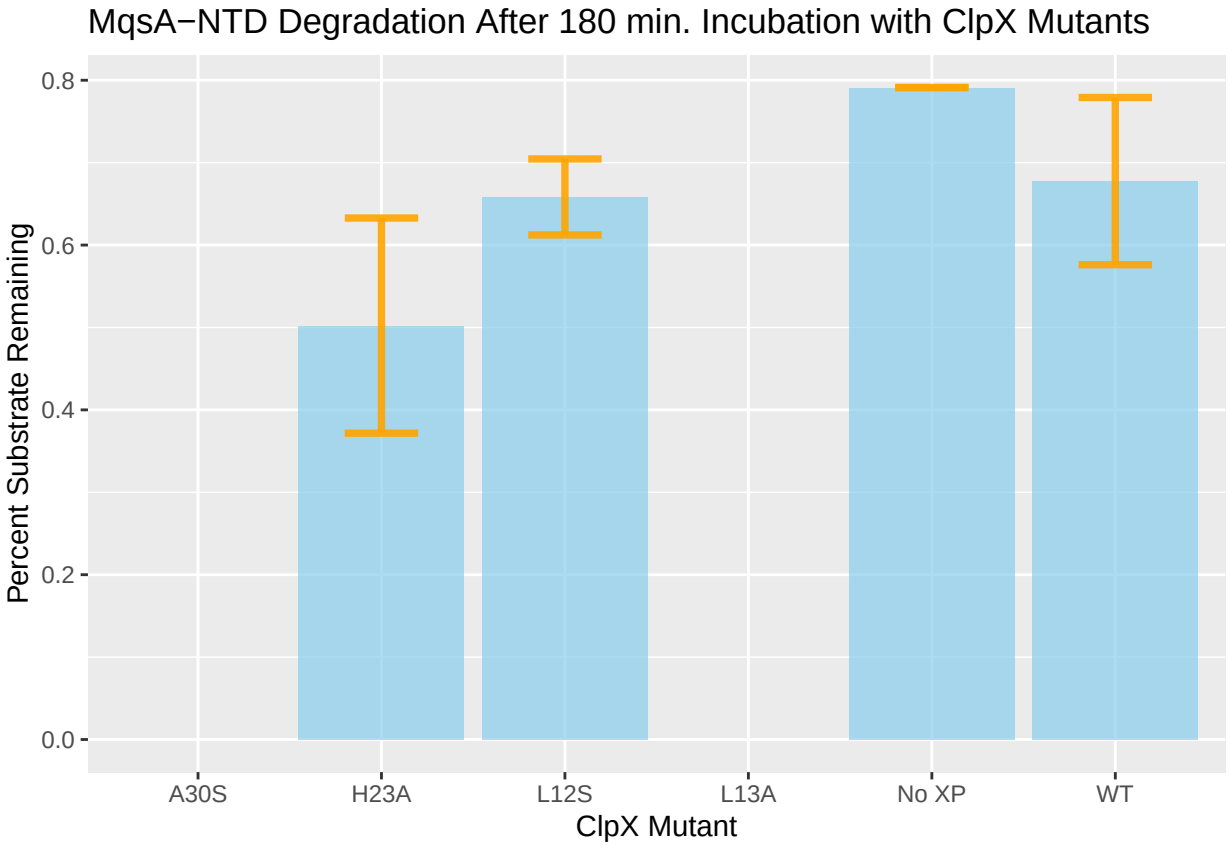
```
ATP_Hydrolysis_Data_for_R_Project_3 %>%
  summary(Avg_Hydrolysis)
```

```
##  ClpX_Mutant        Avg_Hydrolysis         SD
##  Length:6           Min.   :0.0527   Min.   :0.06083
##  Class :character   1st Qu.:0.4467   1st Qu.:0.06627
##  Mode  :character   Median :0.6500   Median :0.09538
##                     Mean   :0.6542   Mean   :0.09339
##                     3rd Qu.:0.9217   3rd Qu.:0.10764
##                     Max.   :1.2000   Max.   :0.13682
##                     NA's   :1        NA's   :1
```

Plot the degradation assays for the ClpX mutants using MqsA-NTD as the substrate

```
ggplot(MqsA_NTD_Degradation_R_Project_2) +
  geom_bar( aes(ClpX_Mutant,MqsA_NTD_Degradation), stat="identity", fill="skyblue", alpha=0.7) +
  geom_errorbar( aes(x=ClpX_Mutant, ymin=MqsA_NTD_Degradation-SD, ymax=MqsA_NTD_Degradation+SD),
                 width=0.4, color="orange", alpha=0.9, size=1.3) +
  labs(title = "MqsA-NTD Degradation After 180 min. Incubation with ClpX Mutants" ,
       x = "ClpX Mutant" , y = "Percent Substrate Remaining")
```

```
## Warning: Removed 2 rows containing missing values (position_stack).
```



#This plots the ClpX mutants on the x-axis and the percent of the MqsA-NTD substrate remaining after 180 minutes on the y-axis. Refer to the plotting of the fluorescence data to see what the other pieces of code here represent.

#The results of this plot show that H23A has highest degradation rate out of all the ClpX mutants. This is interesting because H23A also had a high ATP hydrolysis rate and ATP hydrolysis is directly needed in order for ClpX to unfold its substrates for degradation by ClpP so the beginnings of a correlation are shown here. Unfortunately, A30S and L13A have not been tested in this assay yet but the data set and plot will be updated when those assays are run.

Calculate the summary of the percent of substrate remaining for all ClpX mutants

```
MqsA_NTD_Degradation_R_Project_2 %>%
  summary(MqsA_NTD_Degradation)
```

```
##  ClpX_Mutant        MqsA_NTD_Degradation       SD
##  Length:6           Min.   :0.5023       Min.   :0.00000
##  Class :character   1st Qu.:0.6194       1st Qu.:0.03456
##  Mode  :character   Median :0.6680       Median :0.07377
##                     Mean   :0.6574       Mean   :0.06952
##                     3rd Qu.:0.7060       3rd Qu.:0.10872
##                     Max.   :0.7912       Max.   :0.13053
##                     NA's   :2            NA's   :2
```

Combine the fluorescence data set and the ATP hydrolysis data set

```
ATP_vs_Binding <- left_join(ATP_Hydrolysis_Data_for_R_Project_3, Binding_Data_For_R_Project_4,
                            by = c("ClpX_Mutant","ClpX_Mutant"))
```

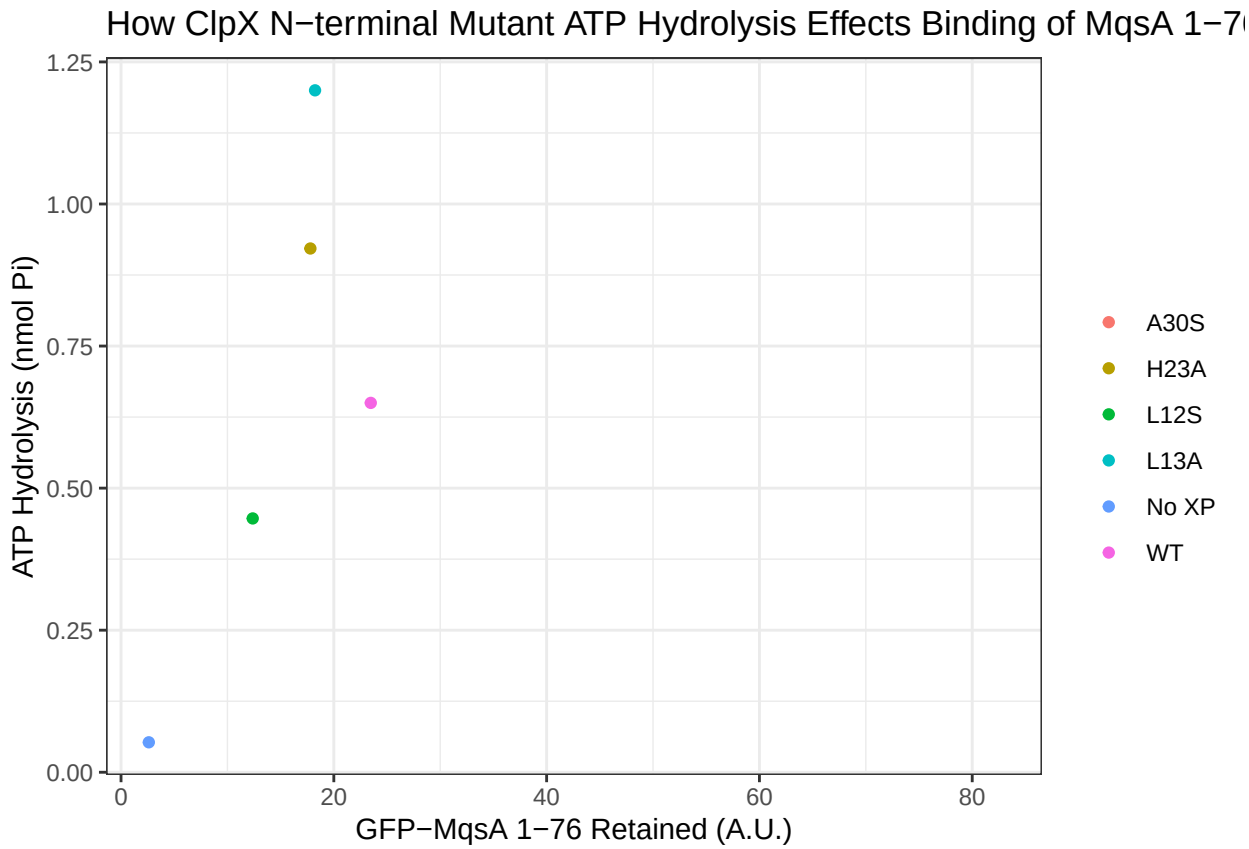
#This has joined the two data sets by the common denominator column, "ClpX_Mutant"

#This provides a new dataset with 5 columns, one for the ClpX_Mutant joining column, then one for each of the recorded data for the respective assays and one for each of their respective standard deviations

Plot the combined ATP-Binding data set

```
ggplot(ATP_vs_Binding) +
  geom_point( aes(MqsA_1_76_GFP_Retained,Avg_Hydrolysis, color = ClpX_Mutant)) +
  theme_bw() +
  labs(title = "How ClpX N-terminal Mutant ATP Hydrolysis Effects Binding of MqsA 1-76" ,
       x = "GFP-MqsA 1-76 Retained (A.U.)" , y = "ATP Hydrolysis (nmol Pi)", color = " ") +
  theme(legend.position = "right")
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



#This script plots the newly combined data set by putting the data from the fluorescence assay on the x-axis and the data from the ATP hydrolysis assay on the y-axis with the ClpX_Mutant column denoting the color and therefore allowing determination of which sample is which.

#This plot is not overly informative because of the fact that ATP hydrolysis and binding of substrates does not have any correlation, however, it is worth observing to see any possible trends of increased or decreased activity in both areas. From this plot, it does not look like there is any correlation to the amount of ATP hydrolyzed and the amount of GFP-MqsA 1-76 bound by any of the ClpX mutants.

Combine the fluorescence data set and the degradation data set

```
Degradation_vs_Binding <- left_join(MqsA_NTD_Degradation_R_Project_2, Binding_Data_For_R_Project_4,
                            by = c("ClpX_Mutant","ClpX_Mutant"))
```

#This has joined the two data sets by the common denominator column, "ClpX_Mutant"

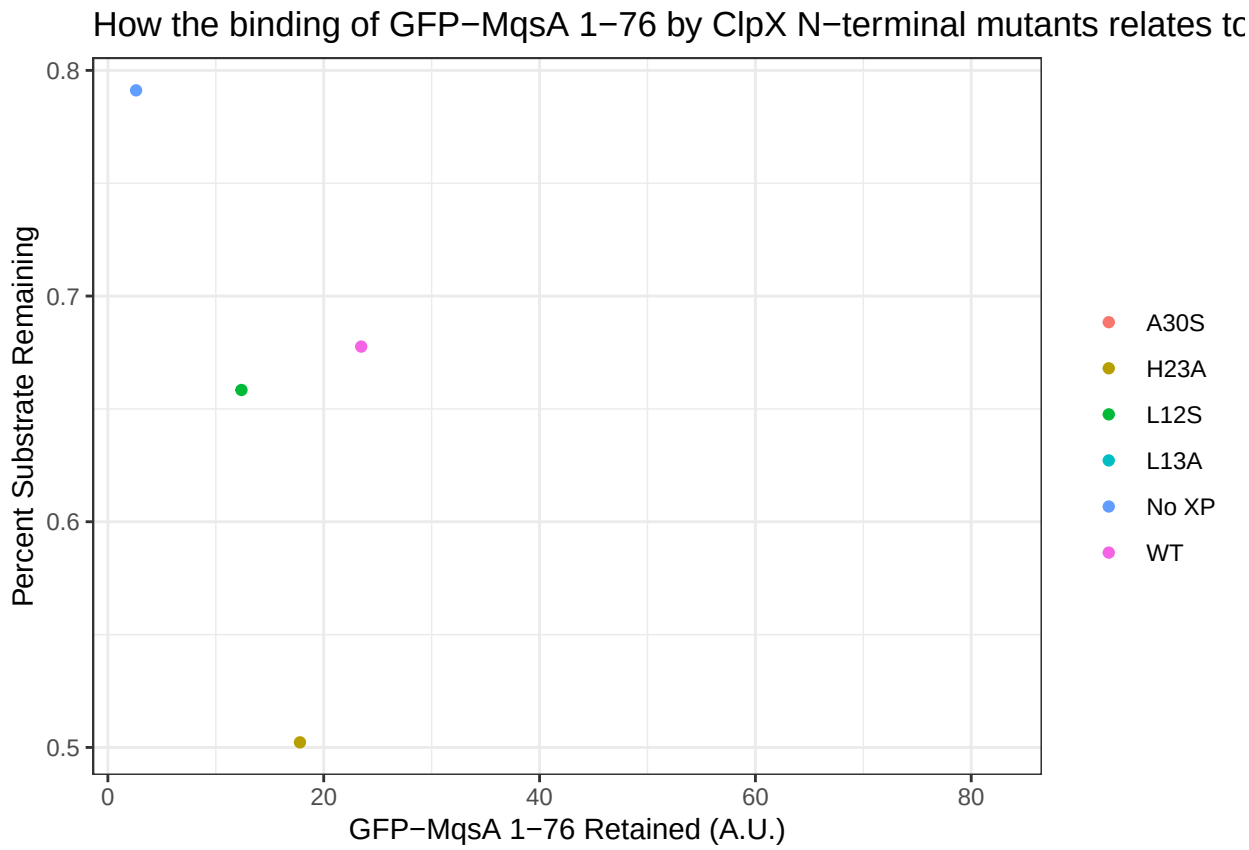
#This provides a new dataset with 5 columns, one for the ClpX_Mutant joining column, then one for each of the recorded data for the respective assays and one for each of their respective standard deviations

Plot the combined Degradation-binding data set

```
ggplot(Degradation_vs_Binding) +
  geom_point( aes(MqsA_1_76_GFP_Retained,MqsA_NTD_Degradation, color = ClpX_Mutant)) +
```

```
  theme_bw() +
  labs(title = "How the binding of GFP-MqsA 1-76 by ClpX N-terminal mutants relates to MqsA-NTD degrada
       x = "GFP-MqsA 1-76 Retained (A.U.)" , y = "Percent Substrate Remaining", color = " ") +
  theme(legend.position = "right")
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



#This script plots the newly combined data set by putting the data from the fluorescence assay on the x-axis and the data from the MqsA-NTD degradation assay on the y-axis with the ClpX_Mutant column denoting the color and therefore allowing determination of which sample is which.

#This plot is particularly interesting because the amount of GFP-MqsA 1-76 retained by the ClpX variant should directly correlate with the amount of MqsA-NTD that is degraded, assuming that the specific mutation in the ClpX protein does not effect the unfolding and transfer of a bound substrate. From the plot, it appears that the WT ClpX has the best degradation of MqsA-NTD while still achieving a high binding affinity for GFP-MqsA 1-76. The tested mutations appear to have some defect in the steps going from binding the substrate to unfolding and transfering it to ClpP for degradation.

#It is worth noting that although the mutations, A30S and L13A were tested in the fluorescence assay, they were not tested in the degradation assay and therefore do not appear on this plot. Once they are tested in that assay, the dataset and plot will be updated

Combine degradation dataset and the ATP hydrolysis data set

```
Degradation_vs_Hydrolysis <- left_join(MqsA_NTD_Degradation_R_Project_2, ATP_Hydrolysis_Data_for_R_Proj
                                       by = c("ClpX_Mutant", "ClpX_Mutant"))
```

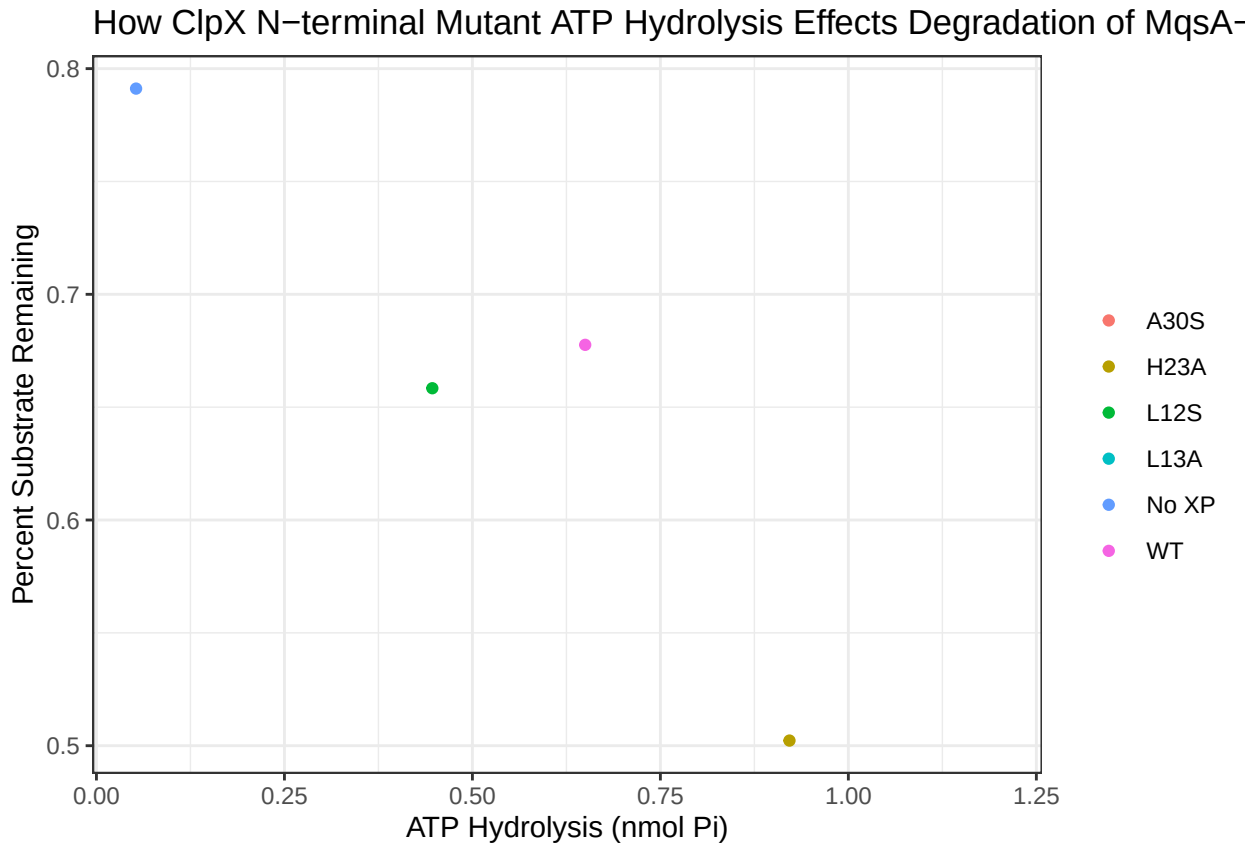
#This has joined the two data sets by the common denominator column, "ClpX_Mutant"

#This provides a new dataset with 5 columns, one for the ClpX_Mutant joining column, then one for each of the recorded data for the respective assays and one for each of their respective standard deviations

Plot the combined degradation-ATP hydrolysis data set

```
ggplot(Degradation_vs_Hydrolysis) +
  geom_point( aes(Avg_Hydrolysis,MqsA_NTD_Degradation, color = ClpX_Mutant)) +
  theme_bw() +
  labs(title = "How ClpX N-terminal Mutant ATP Hydrolysis Effects Degradation of MqsA-NTD" ,
       x = "ATP Hydrolysis (nmol Pi)" , y = "Percent Substrate Remaining", color = " ") +
  theme(legend.position = "right")
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



#This script plots the newly combined data set by putting the data from the ATP hydrolysis assay on the x-axis and the data from the MqsA-NTD degradation assay on the y-axis with the ClpX_Mutant column denoting the color and therefore allowing determination of which sample is which.

#The results of this plot show that H23A has the largest correlation between ATP hydrolysis rate and MqsA-NTD degradation. It has the highest hydrolysis rate and the most degradation of the substrate. ClpX utilizes ATP to power the unfolding of substrates to then be transfered to ClpP for degradation so this correlation makes sense in terms of proteolytic mechanisms. Again, A30S and L13A were not tested in both assays so they are ommitted from this plot. When those assays are run, the data set and plot will be updated to reflect those results.