20210608_Drp1_variants_analysis.R

kelse

2022-06-01

```r
# Script for analyzing and visualizing SEC-MALS data
# Data collected on 20210608
# Drp1 constructs dialyzed overnight into SEC-MALS buffer
# Exact buffer match to running buffer used
# 6xHis tag present


library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.0      v dplyr   1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## Warning: package 'stringr' was built under R version 4.0.5
```

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

```r
library(dplyr)
library(readxl)
library(caret)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(RColorBrewer)
library(broom)
library(minpack.lm)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##      combine
```

```
theme_set(theme_bw() +
            theme(axis.text = element_text(size = 12, color = "black"),
                  panel.grid.major = element_blank(),
                  panel.grid.minor = element_blank()))

# Import excel files containing raw data output from ASTRA software and tidy data.
WT_raw <- read_xlsx("20210608_Drp1_variants_analysis.xlsx",
                          sheet = 8) %>%
 gather(., key = "tmp1", value = "intensity", 2:21) %>%
  separate(., col = "tmp1", into = c("prot", "detector"), sep = "_")

G363D_raw <- read_xlsx("20210608_Drp1_variants_analysis.xlsx",
                        sheet = 7) %>%
  gather(., key = "tmp1", value = "intensity", 2:21) %>%
  separate(., col = "tmp1", into = c("prot", "detector"), sep = "_")

G401S_raw <- read_xlsx("20210608_Drp1_variants_analysis.xlsx",
                        sheet = 6) %>%
  gather(., key = "tmp1", value = "intensity", 2:21) %>%
  separate(., col = "tmp1", into = c("prot", "detector"), sep = "_")

R710G_raw <- read_xlsx("20210608_Drp1_variants_analysis.xlsx",
                        sheet = 5) %>%
  gather(., key = "tmp1", value = "intensity", 2:21) %>%
  separate(., col = "tmp1", into = c("prot", "detector"), sep = "_")

# Import excel files containing MW data output from ASTRA software and tidy data.
WT_mm <- read_xlsx("20210608_Drp1_variants_analysis.xlsx",
                        sheet = 4) %>%
  gather(., key = "tmp1", value = "mw", 2:4) %>%
  separate(., col = "tmp1", into = c("prot", "peak"), sep = "_")

G363D_mm <- read_xlsx("20210608_Drp1_variants_analysis.xlsx",
                        sheet = 3) %>%
  gather(., key = "tmp1", value = "mw", 2:4) %>%
  separate(., col = "tmp1", into = c("prot", "peak"), sep = "_")

G401S_mm <- read_xlsx("20210608_Drp1_variants_analysis.xlsx",
                        sheet = 2) %>%
  gather(., key = "tmp1", value = "mw", 2:3) %>%
  separate(., col = "tmp1", into = c("prot", "peak"), sep = "_")

R710G_mm <- read_xlsx("20210608_Drp1_variants_analysis.xlsx",
                        sheet = 1) %>%
  gather(., key = "tmp1", value = "mw", 2:3) %>%
  separate(., col = "tmp1", into = c("prot", "peak"), sep = "_")
```

```r
### Merge data sets into one tidy table ###
data <- WT_raw %>%
  union(., R710G_raw) %>%
  union(., G401S_raw) %>%
  union(., G363D_raw)

mw_data <- WT_mm %>%
  union(., G363D_mm) %>%
  union(., G401S_mm) %>%
  union(., R710G_mm)

### Normalize data ###
normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}

# omit rows with NA #
data_no_na <- na.omit(data)
mw_no_na <- na.omit(mw_data)

# truncate subset of dRI data to only show times <= 12 minutes
data_dRI_trunc <- data_no_na[which(data_no_na$detector=='dRI'
                        & data_no_na$time <= 12), ]

# truncate subset of 'UV data to only show times <= 12 minutes
data_trunc_UV <- data_no_na[which(data_no_na$detector=='UV'
                             & data_no_na$time <= 15), ]

# truncate subset of LS9 data to only show times <= 12 minutes
data_trunc_LS9 <- data_no_na[which(data_no_na$detector=='LS9'
                             & data_no_na$time <= 12), ]

#truncate all data to only show times <=12 minutes
data_trunc <- data_no_na[which(data_no_na$time <=12), ]

#Normalize intensities of truncated dRI data
data_dRI_trunc$norm_intensity <- normalize(data_dRI_trunc$intensity)

#Normalize intensities of truncated UV data
data_trunc_UV$norm_intensity <- normalize(data_trunc_UV$intensity)

#Normalize intensities of truncated LS9 data
data_trunc_LS9$norm_intensity <- normalize(data_trunc_LS9$intensity)

#Normalize intensities of truncated data (all)
data_trunc$norm_intensity <- normalize(data_trunc$intensity)

# centering with 'scale()'
center_scale <- function(x) {
  scale(x, scale = FALSE)
}

# apply scaling to normalized truncated dRI
```

```r
data_dRI_trunc$scaled_norm_intensity <- center_scale(data_dRI_trunc$norm_intensity)

# apply scaling to normalized truncated UV
data_trunc_UV$scaled_norm_intensity <- center_scale(data_trunc_UV$norm_intensity)

# apply scaling to normalized truncated LS9
data_trunc_LS9$scaled_norm_intensity <- center_scale(data_trunc_LS9$norm_intensity)

# apply scaling to normalized truncated data
data_trunc$scaled_norm_intensity <- center_scale(data_trunc$norm_intensity)

# select only dRI data
data_dRI <- data_no_na[which(data_no_na$detector=='dRI'), ]

#Normalize intensities of dRI data
data_dRI$norm_intensity <- normalize(data_dRI$intensity)

# apply scaling to normalized dRI
data_dRI$scaled_norm_intensity <- center_scale(data_dRI$norm_intensity)

#normalize full data set
data$norm_intensity <- normalize(data$intensity)

# apply scaling to normalized full data set
data$scaled_norm_intensity <- center_scale(data$norm_intensity)

# Plot calculated molecular weight values
mw_data %>%
  mutate(prot =
           factor(prot,
                  levels = c("WT", "L230dup", "G363D", "G401S", "R710G"))) %>%
  ggplot(., aes(x=time, y = as.numeric(mw), color = prot)) +
  geom_point(size = 0.1) +
  scale_color_manual(values = c("WT" = "#7570B3",
                                "L230dup" = "#66A61E",
                                "G363D" = "#1B9E77",
                                "G401S" = "#D95F02",
                                "R710G" = "#E7298A")) +
    scale_x_continuous(limits = c(4,15),
                       breaks = c(4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)) +
    scale_y_continuous(limits = c(0, 480000),
                       breaks = c(0, 80000, 160000, 240000, 320000, 400000, 480000)) +
  labs(title = "MW: Drp1 and Variants",
       x = "Elution time (min)",
       y = "MW",
       color = "Protein") +
  theme(plot.title = element_text (hjust = 0.5))
```

```
## Warning: Removed 1468 rows containing missing values (geom_point).
```
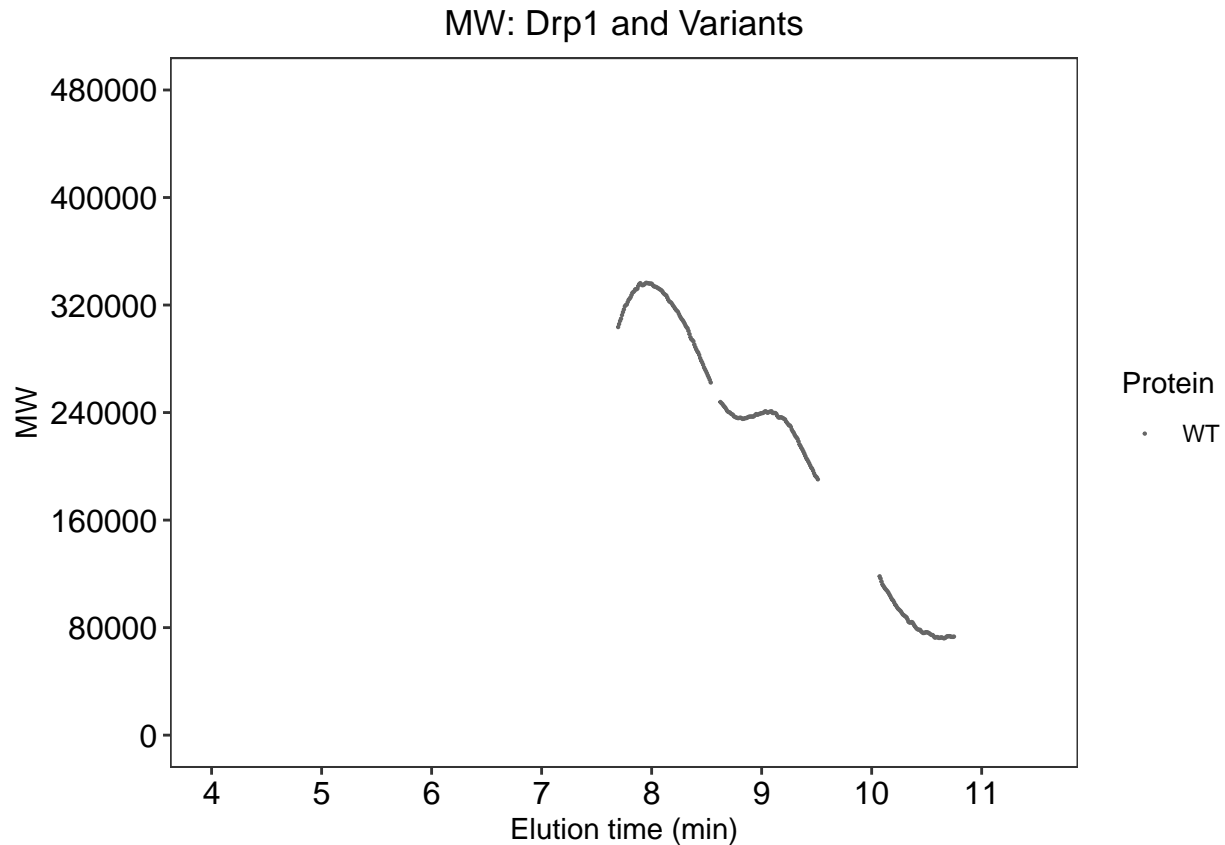
MW: Drp1 and Variants
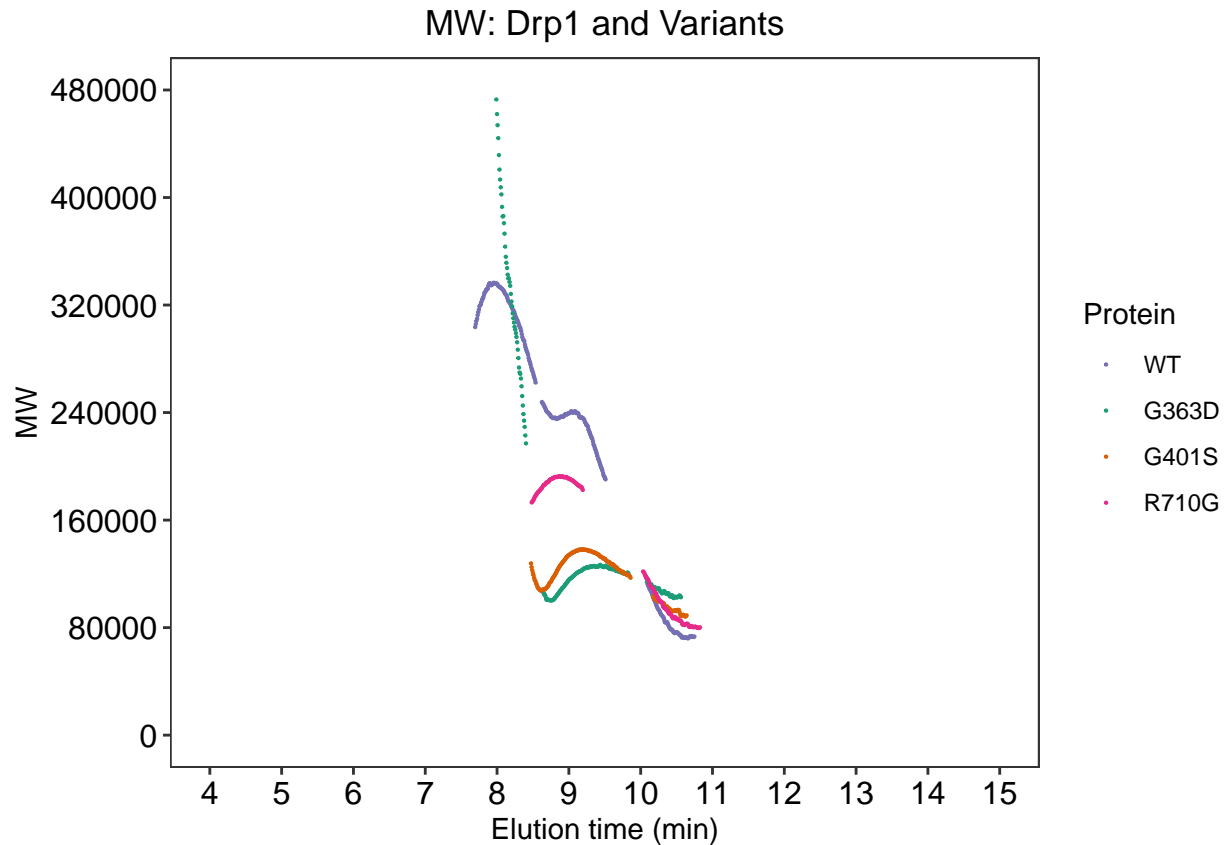
```
ggsave("MW_calcs_trunc_allsamples.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```

Warning: Removed 1468 rows containing missing values (geom_point).

```
# Plot calculated molecular weight values just WT Drp1
WT_MW <- mw_data %>%
  filter(., prot == "WT") %>%
    ggplot(., aes(x=time, y = as.numeric(mw), color = prot)) +
  geom_point(size = 0.1) +
  scale_color_manual(values = c("WT" = "gray40")) +
  scale_x_continuous(limits = c(4,11.5),
                     breaks = c(4, 5, 6, 7, 8, 9, 10, 11)) +
  scale_y_continuous(limits = c(0, 480000),
                     breaks = c(0, 80000, 160000, 240000, 320000, 400000, 480000)) +
  labs(title = "MW: Drp1 and Variants",
       x = "Elution time (min)",
       y = "MW",
       color = "Protein") +
  theme(plot.title = element_text (hjust = 0.5))

WT_MW
```

Warning: Removed 554 rows containing missing values (geom_point).

MW: Drp1 and Variants


```
ggsave("MW_calcs_trunc_WT.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```
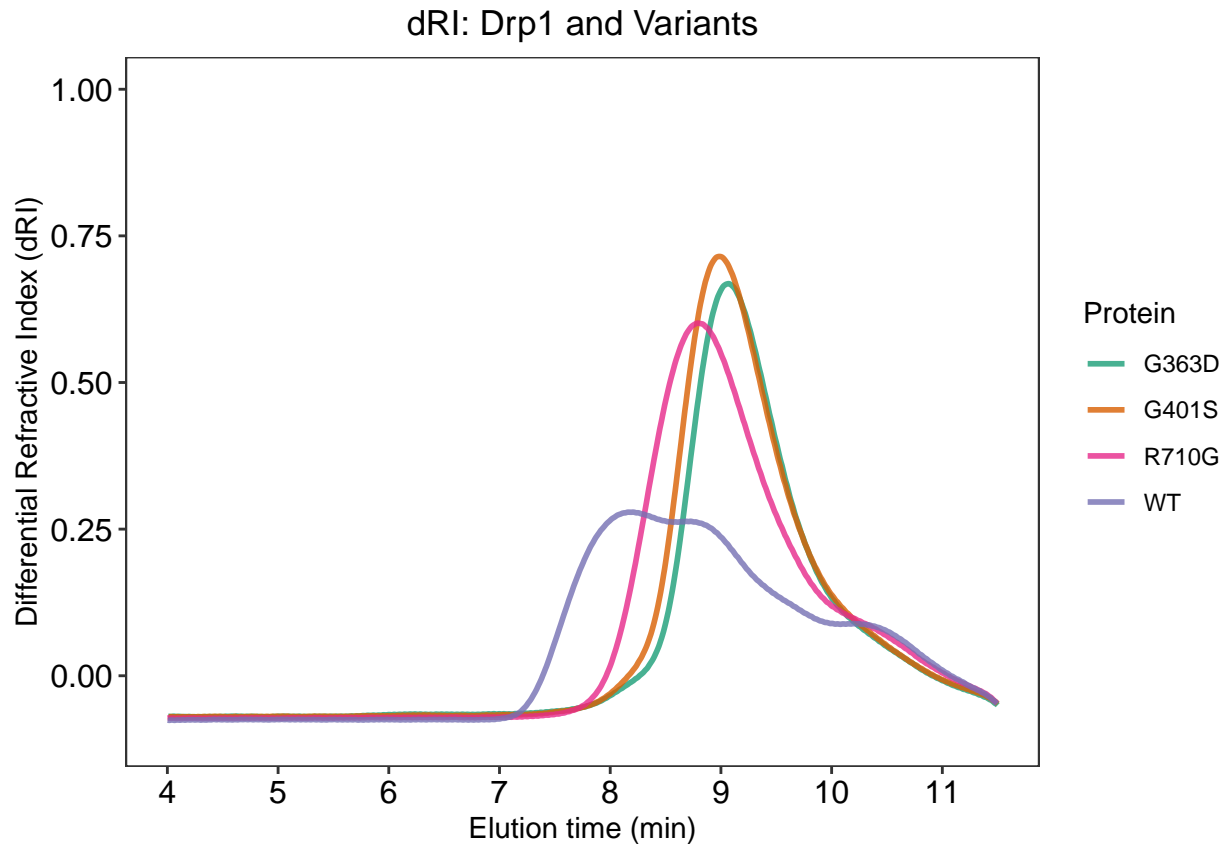
```
## Warning: Removed 554 rows containing missing values (geom_point).
```

```
#Plot normalized and scaled truncated dRI overlays of all samples

data_dRI_trunc %>%
  mutate(prot =
           factor(prot,
                  levels = c("WT", "G363D", "G401S", "R710G"))) %>%
  ggplot(., aes(x = time)) +
  geom_line(data = data_dRI_trunc, aes(y = scaled_norm_intensity, color = prot), size = 1, alpha = 0.8)
  scale_color_manual(values = c("WT" = "#7570B3",
                                "G363D" = "#1B9E77",
                                "G401S" = "#D95F02",
                                "R710G" = "#E7298A")) +
  scale_x_continuous(limits = c(4,11.5),
                     breaks = c(4, 5, 6, 7, 8, 9, 10, 11)) +
  scale_y_continuous(limits = c(-0.1, 1),
                     breaks = c(-0, 0.25, 0.5, 0.75, 1)) +
  labs(title = "dRI: Drp1 and Variants",
       x = "Elution time (min)",
       y = "Differential Refractive Index (dRI)",
       color = "Protein",
```

```
        shape = "Replicate") +
  theme(plot.title = element_text (hjust = 0.5))
```

Warning: Removed 2398 row(s) containing missing values (geom_path).


```
ggsave("dRI_norm_trunc_allsamples.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```

Warning: Removed 2398 row(s) containing missing values (geom_path).

```
#Plot normalized and scaled truncated dRI overlays of just WT Drp1

WT_dRI <- data_dRI_trunc %>%
    filter(., (prot == "WT")) %>%
  ggplot(., aes(x = time)) +
  geom_line(aes(y = scaled_norm_intensity, color = prot), size = 1, alpha = 0.8) +
  scale_color_manual(values = c("WT" = "#7570B3")) +
  scale_x_continuous(limits = c(4,11),
                     breaks = c(4, 5, 6, 7, 8, 9, 10, 11)) +
  scale_y_continuous(limits = c(-0.1, 1),
                     breaks = c(-0, 0.25, 0.5, 0.75, 1)) +
  labs(title = "dRI: Drp1 and Variants",
       x = "Elution time (min)",
```

```
        y = "Differential Refractive Index (dRI)",
        color = "Protein",
        shape = "Replicate") +
    theme(plot.title = element_text (hjust = 0.5))

ggsave("dRI_norm_trunc_Drp1_WT.pdf",
        width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```

```
## Warning: Removed 656 row(s) containing missing values (geom_path).
```
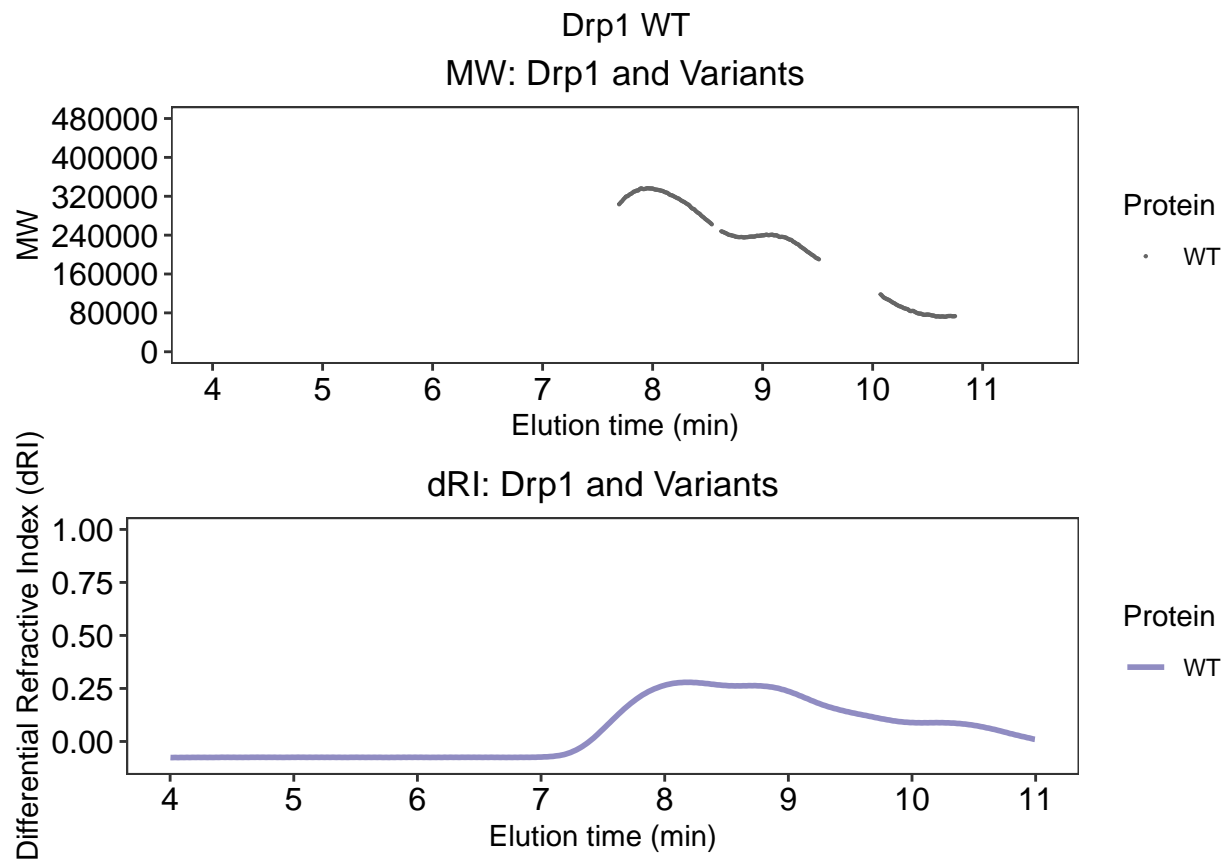
```
# plot WT MW + dRI together
plot1 <- grid.arrange(grobs = list(WT_MW, WT_dRI),
                      top = "Drp1 WT")
```

```
## Warning: Removed 554 rows containing missing values (geom_point).
```

```
## Warning: Removed 656 row(s) containing missing values (geom_path).
```



```
plot1
```

```
## TableGrob (3 x 1) "arrange": 3 grobs
##   z     cells     name                grob
## 1 1 (2-2,1-1) arrange     gtable[layout]
## 2 2 (3-3,1-1) arrange     gtable[layout]
## 3 3 (1-1,1-1) arrange text[GRID.text.536]
```
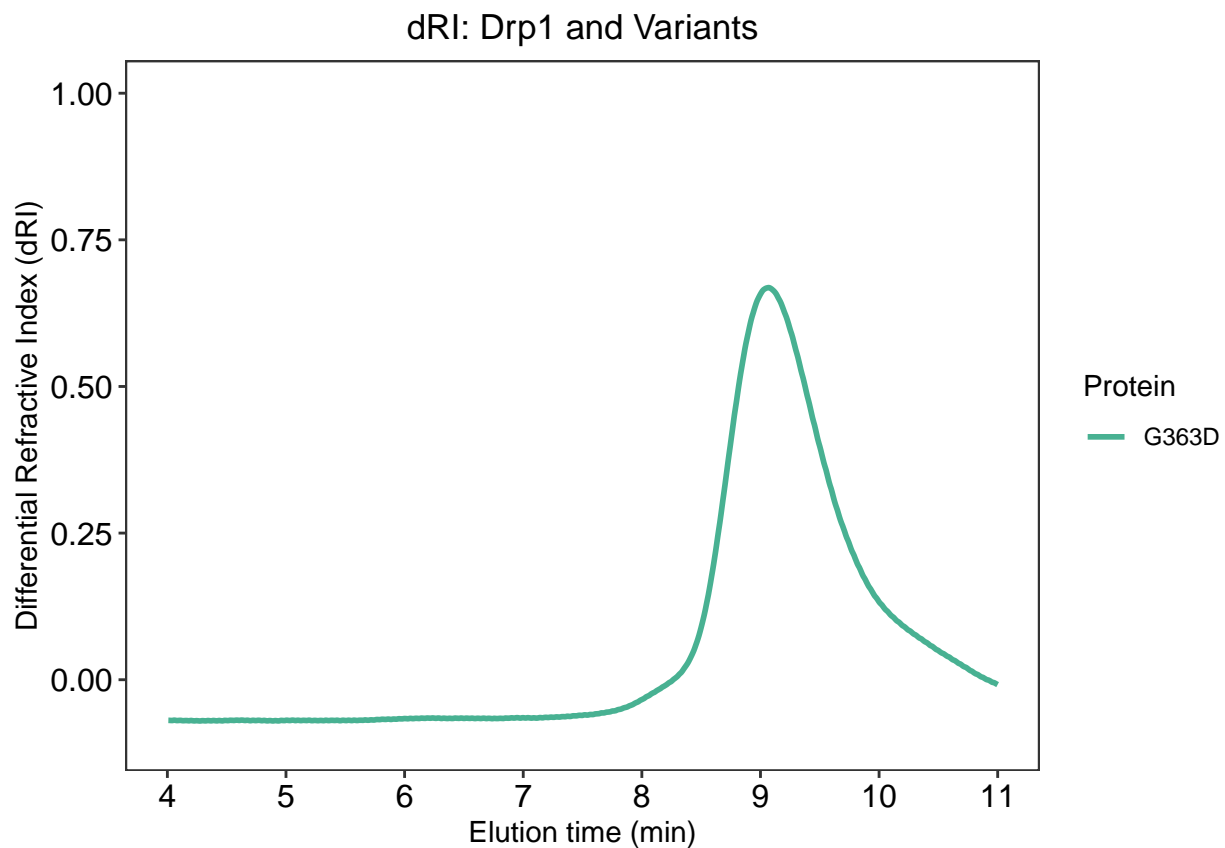
```
ggsave("WT_MW_dRI.pdf", plot1,
       width = 10, height = 12, units = "cm")

#Plot normalized and scaled truncated dRI overlays of just Drp1 G363D
data_dRI_trunc %>%
  filter(., (prot == "G363D")) %>%
  ggplot(., aes(x = time)) +
  geom_line(aes(y = scaled_norm_intensity, color = prot), size = 1, alpha = 0.8) +
  scale_color_manual(values = c("G363D" = "#1B9E77")) +
  scale_x_continuous(limits = c(4,11),
                     breaks = c(4, 5, 6, 7, 8, 9, 10, 11)) +
  scale_y_continuous(limits = c(-0.1, 1),
                     breaks = c(-0, 0.25, 0.5, 0.75, 1)) +
  labs(title = "dRI: Drp1 and Variants",
       x = "Elution time (min)",
       y = "Differential Refractive Index (dRI)",
       color = "Protein",
       shape = "Replicate") +
  theme(plot.title = element_text (hjust = 0.5))
```

```
## Warning: Removed 657 row(s) containing missing values (geom_path).
```
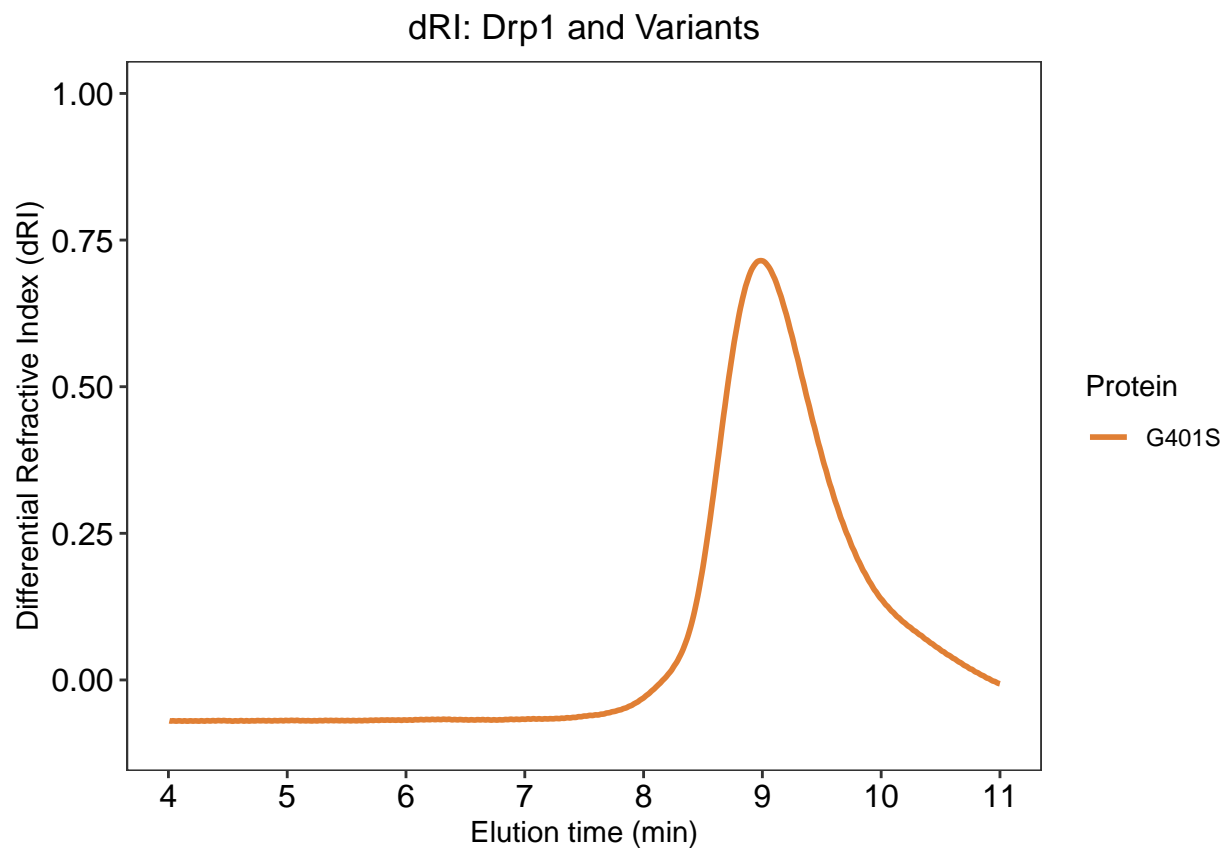


```
ggsave("dRI_norm_trunc_Drp1_G363D.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```

```
## Warning: Removed 657 row(s) containing missing values (geom_path).
```

```
#Plot normalized and scaled truncated dRI overlays of just Drp1 G401S
data_dRI_trunc %>%
  filter(., (prot == "G401S")) %>%
  ggplot(., aes(x = time)) +
  geom_line(aes(y = scaled_norm_intensity, color = prot), size = 1, alpha = 0.8) +
  scale_color_manual(values = c("G401S" = "#D95F02")) +
  scale_x_continuous(limits = c(4,11),
                     breaks = c(4, 5, 6, 7, 8, 9, 10, 11)) +
  scale_y_continuous(limits = c(-0.1, 1),
                     breaks = c(-0, 0.25, 0.5, 0.75, 1)) +
  labs(title = "dRI: Drp1 and Variants",
       x = "Elution time (min)",
       y = "Differential Refractive Index (dRI)",
       color = "Protein",
       shape = "Replicate") +
  theme(plot.title = element_text (hjust = 0.5))
```

```
## Warning: Removed 657 row(s) containing missing values (geom_path).
```
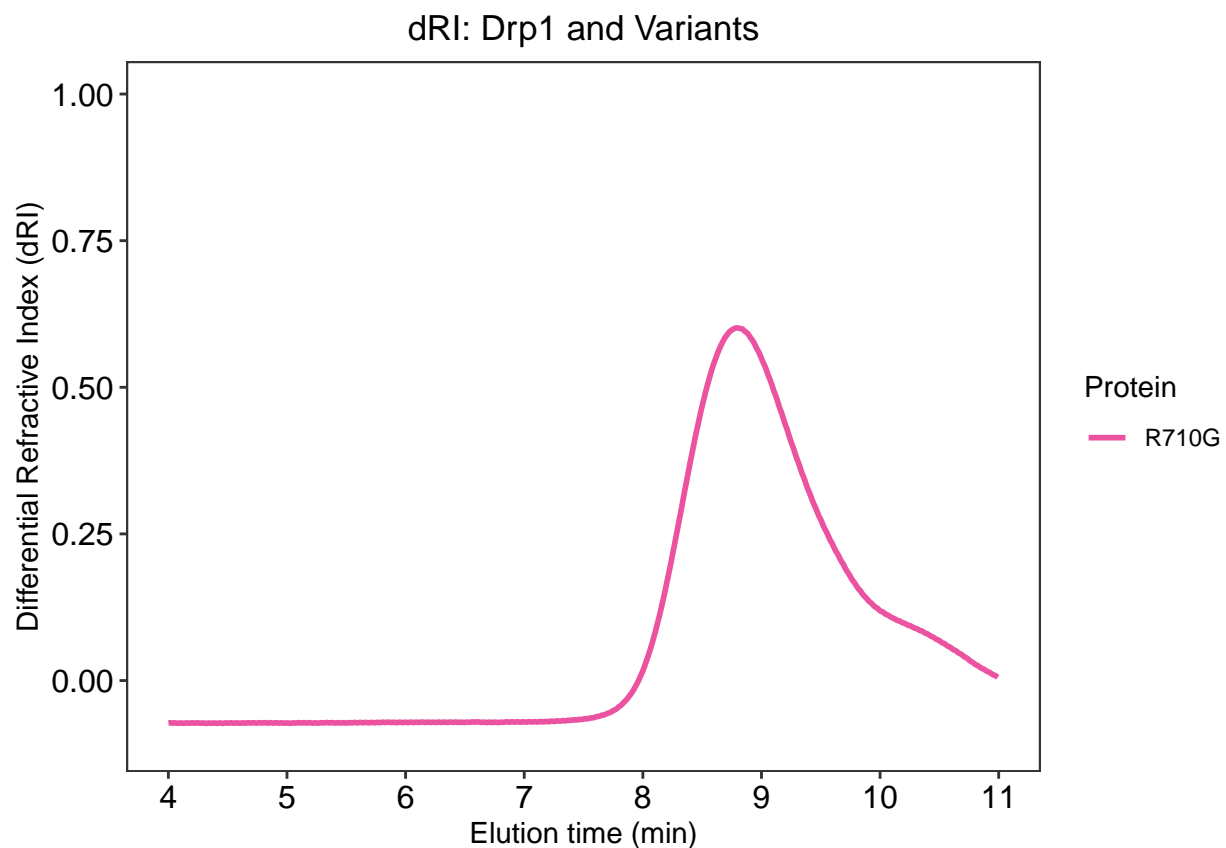


```
ggsave("dRI_norm_trunc_Drp1_G401S.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```

```
## Warning: Removed 657 row(s) containing missing values (geom_path).
```

```
#Plot normalized and scaled truncated dRI overlays of just Drp1 R710G
data_dRI_trunc %>%
  filter(., (prot == "R710G")) %>%
  ggplot(., aes(x = time)) +
  geom_line(aes(y = scaled_norm_intensity, color = prot), size = 1, alpha = 0.8) +
  scale_color_manual(values = c("R710G" = "#E7298A")) +
  scale_x_continuous(limits = c(4,11),
                     breaks = c(4, 5, 6, 7, 8, 9, 10, 11)) +
  scale_y_continuous(limits = c(-0.1, 1),
                     breaks = c(-0, 0.25, 0.5, 0.75, 1)) +
  labs(title = "dRI: Drp1 and Variants",
       x = "Elution time (min)",
       y = "Differential Refractive Index (dRI)",
       color = "Protein",
       shape = "Replicate") +
  theme(plot.title = element_text (hjust = 0.5))
```

Warning: Removed 656 row(s) containing missing values (geom_path).


```
ggsave("dRI_norm_trunc_Drp1_R710G.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```

Warning: Removed 656 row(s) containing missing values (geom_path).

```
#Plot dRI overlays of all samples (full spectra)
data_dRI %>%
  ggplot(., aes(x = time)) +
  geom_line(aes(y = scaled_norm_intensity, color = prot), size = 1, alpha = 0.8) +
  scale_color_brewer(palette="Set2") +
  scale_y_continuous(limits = c(-0.2, 1),
                     breaks = c(0, 0.25, 0.5, 0.75, 1.0)) +
  labs(title = "dRI: Drp1 and Variants",
       x = "Elution time (min)",
       y = "Differential Refractive Index (dRI)",
       color = "Protein",
       shape = "Replicate") +
  theme(plot.title = element_text (hjust = 0.5))
```
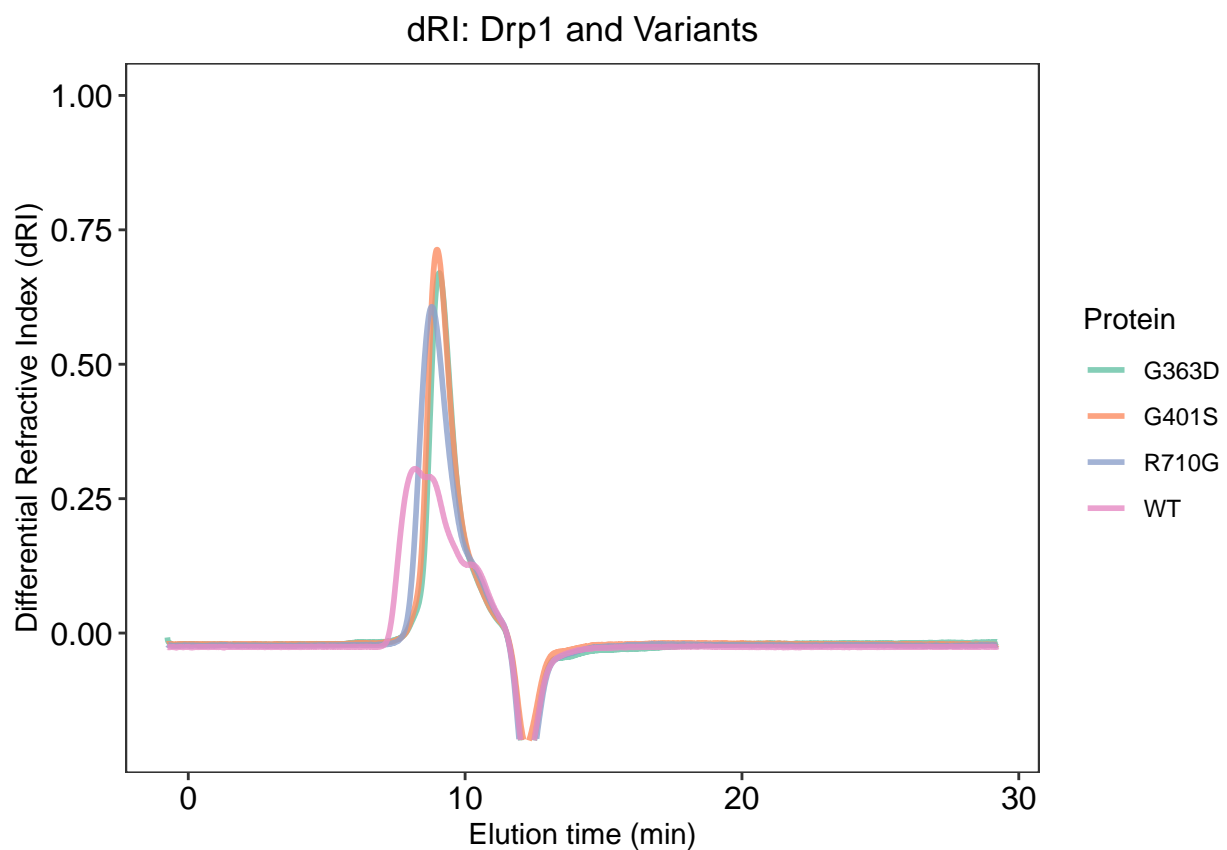


```
ggsave("dRI_allsamples_fullspectra.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)


# Plot LS overlays of all samples from LS detector #9
data_trunc_LS9 %>%
  ggplot(., aes(x = time)) +
  geom_line(aes(y = scaled_norm_intensity, color = prot), size = 1, alpha = 0.8) +
  scale_color_brewer(palette="Set2") +
  scale_x_continuous(limits = c(4,11),
                     breaks = c(4, 5, 6, 7, 8, 9, 10, 11)) +
```
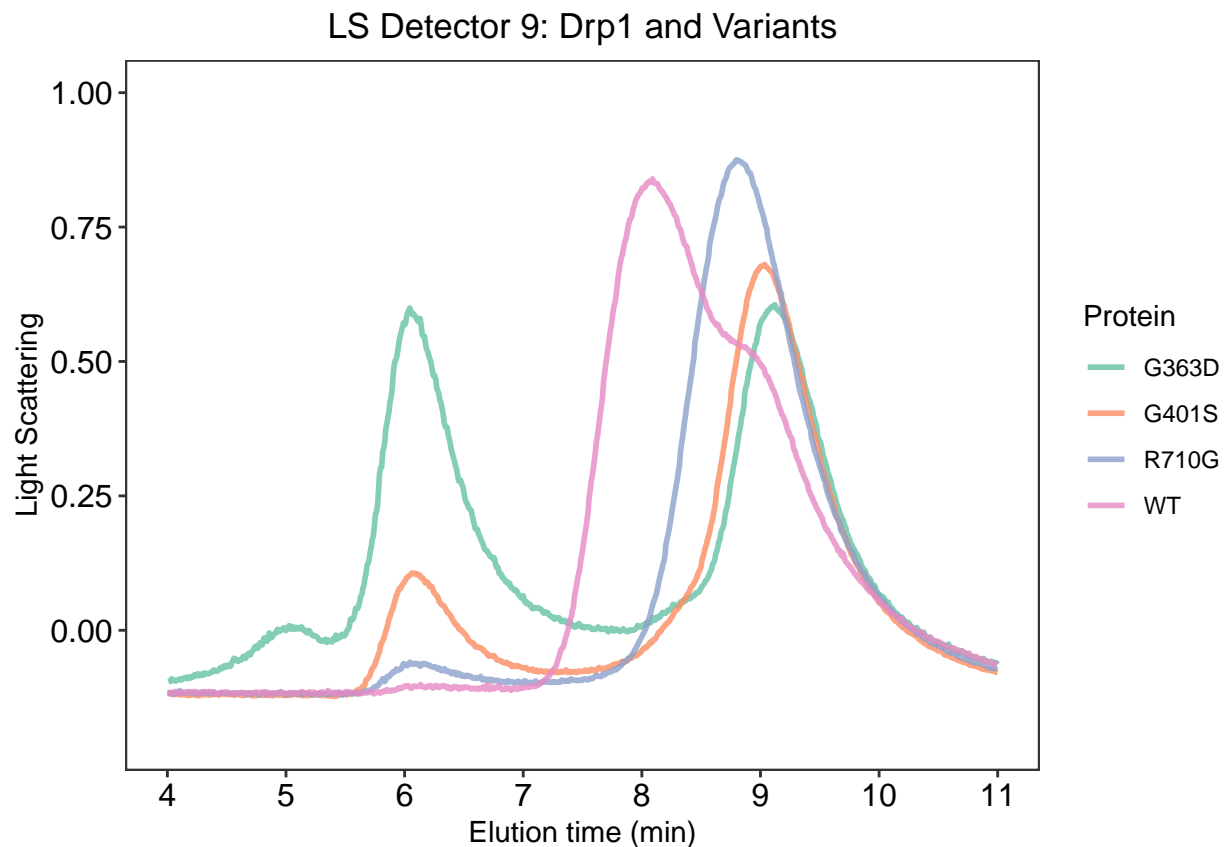
```
    scale_y_continuous(limits = c(-0.2, 1),
                        breaks = c(0, 0.25, 0.5, 0.75, 1.0)) +
    labs(title = "LS Detector 9: Drp1 and Variants",
         x = "Elution time (min)",
         y = "Light Scattering",
         color = "Protein") +
    theme(plot.title = element_text (hjust = 0.5))
```

```
## Warning: Removed 2459 row(s) containing missing values (geom_path).
```



```
ggsave("LS9_allsamples_trunc.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```

```
## Warning: Removed 2459 row(s) containing missing values (geom_path).
```

```
# Plot LS overlays of just Drp1 from LS detector #9
data_trunc_LS9 %>%
  filter(., prot == "WT") %>%
  ggplot(., aes(x = time)) +
  geom_line(aes(y = scaled_norm_intensity, color = prot), size = 1, alpha = 0.8) +
  scale_color_brewer(palette="Set2") +
  scale_x_continuous(limits = c(4,11),
                     breaks = c(4, 5, 6, 7, 8, 9, 10, 11)) +
```
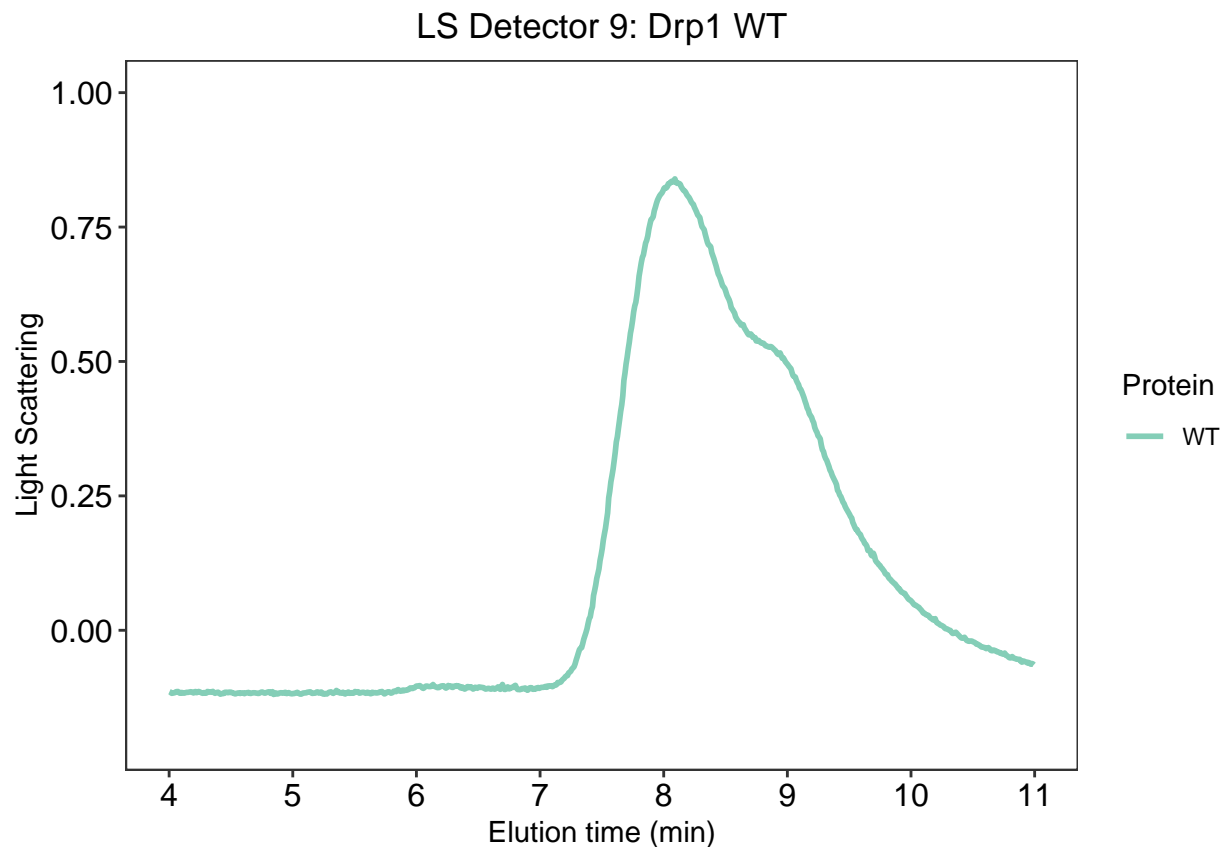
```
    scale_y_continuous(limits = c(-0.2, 1),
                       breaks = c(0, 0.25, 0.5, 0.75, 1.0)) +
    labs(title = "LS Detector 9: Drp1 WT",
         x = "Elution time (min)",
         y = "Light Scattering",
         color = "Protein") +
    theme(plot.title = element_text (hjust = 0.5))
```

Warning: Removed 615 row(s) containing missing values (geom_path).


```
ggsave("LS9_WT_trunc.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```

Warning: Removed 615 row(s) containing missing values (geom_path).

```
# Plot UV chromatograms of all samples
data_trunc_UV %>%
  ggplot(., aes(x = time)) +
  geom_line(aes(y = scaled_norm_intensity, color = prot), size = 1, alpha = 0.8) +
  scale_color_manual(breaks = c("WT", "G363D", "G401S", "R710G"),
                     values = c("WT" = "red",
                               "G363D" = "forestgreen",
                               "G401S" = "purple",
```
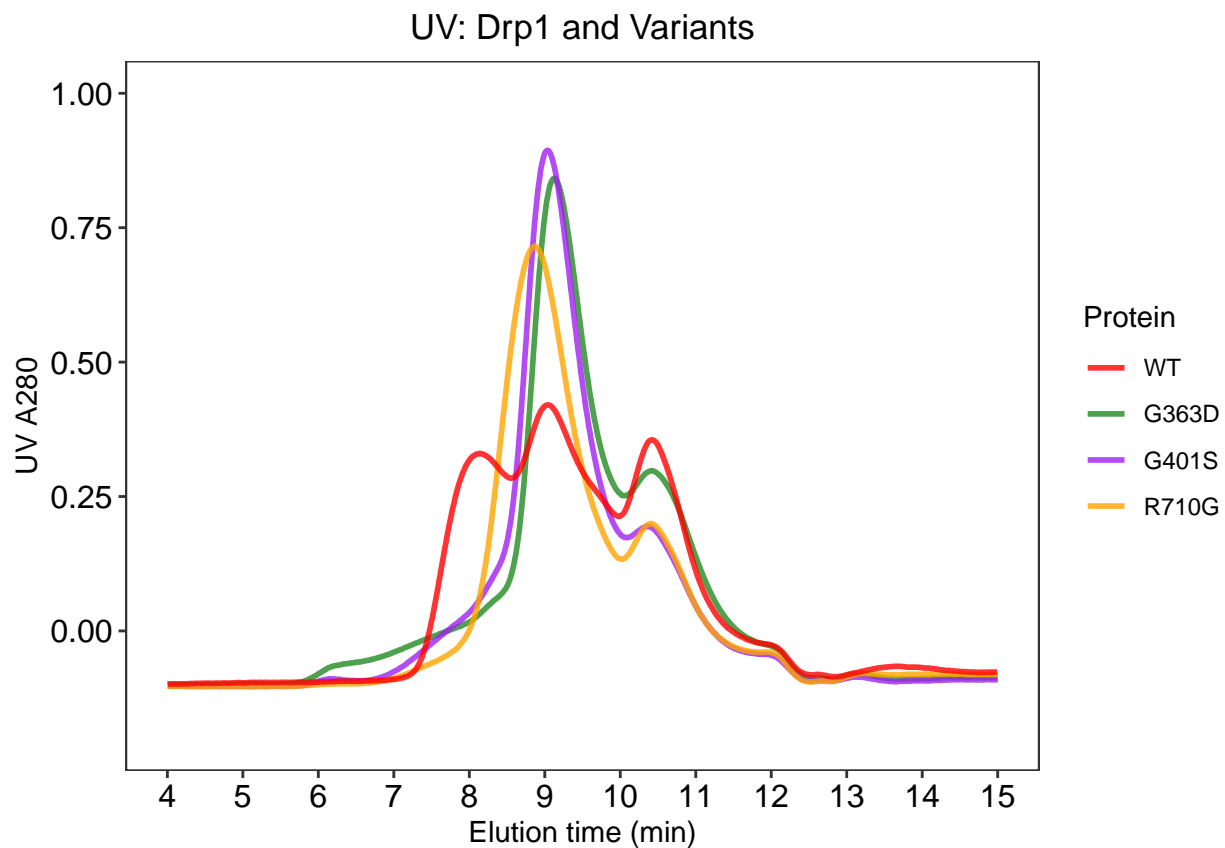
```
                                  "R710G" = "orange")) +
  scale_x_continuous(limits = c(4,15),
                     breaks = c(4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15)) +
  scale_y_continuous(limits = c(-0.2, 1),
                     breaks = c(0, 0.25, 0.5, 0.75, 1.0)) +
  labs(title = "UV: Drp1 and Variants",
       x = "Elution time (min)",
       y = "UV A280",
       color = "Protein") +
  theme(plot.title = element_text (hjust = 0.5))
```

`## Warning: Removed 1824 row(s) containing missing values (geom_path).`


```
ggsave("UV_allsamples_trunc_noL230dup.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```

`## Warning: Removed 1824 row(s) containing missing values (geom_path).`

```
# Plot UV chromatograms of just Drp1 WT
data_trunc_UV %>%
  filter(., prot == "WT") %>%
  ggplot(., aes(x = time)) +
  geom_line(aes(y = scaled_norm_intensity, color = prot), size = 1, alpha = 0.8) +
```
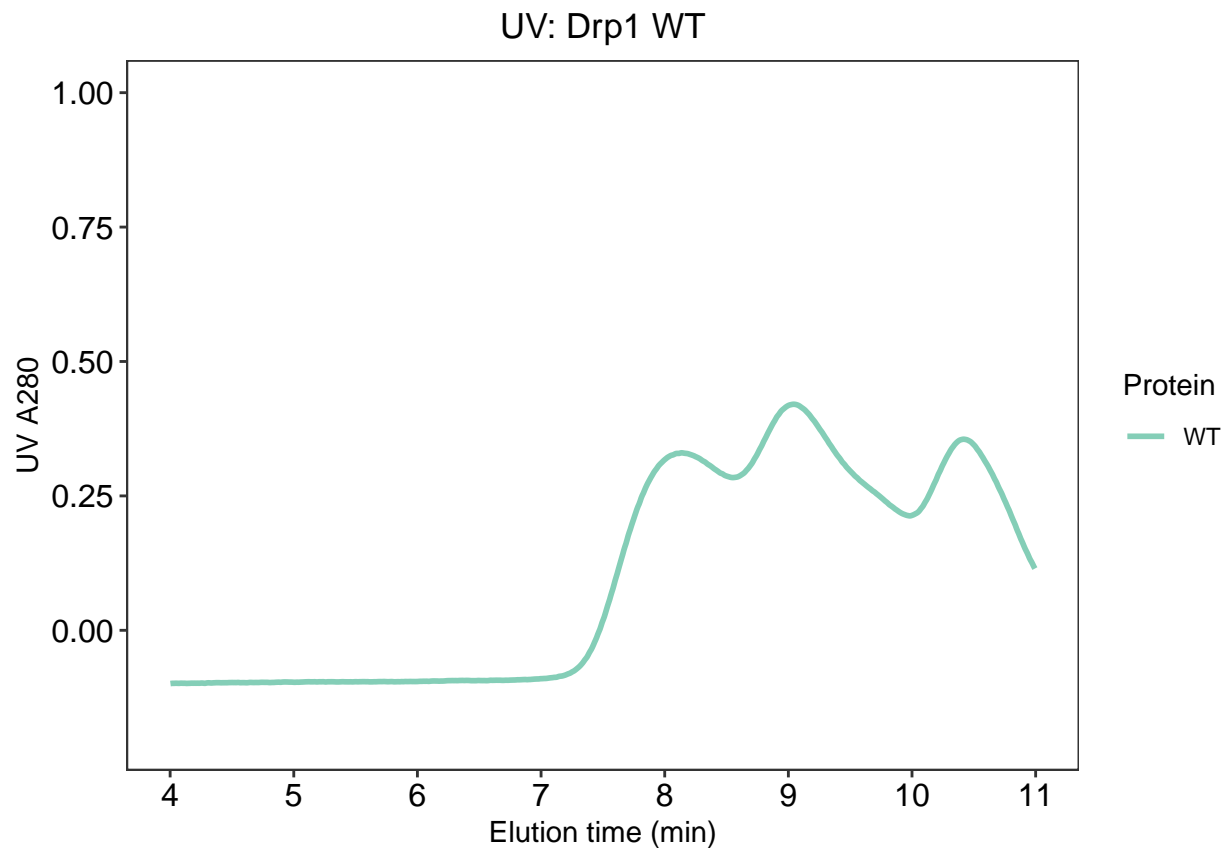
```
    scale_color_brewer(palette="Set2") +
    scale_x_continuous(limits = c(4,11),
                       breaks = c(4, 5, 6, 7, 8, 9, 10, 11)) +
    scale_y_continuous(limits = c(-0.2, 1),
                       breaks = c(0, 0.25, 0.5, 0.75, 1.0)) +
    labs(title = "UV: Drp1 WT",
         x = "Elution time (min)",
         y = "UV A280",
         color = "Protein") +
    theme(plot.title = element_text (hjust = 0.5))
```

Warning: Removed 912 row(s) containing missing values (geom_path).


```
ggsave("UV_Drp1_trunc.pdf",
       width = 6, height = 3, dpi = 300, useDingbats = FALSE)
```

Warning: Removed 912 row(s) containing missing values (geom_path).

```
#compute time of max signal dRI
# output = value of max scaled and normalized intensity
# need to parse for time
data_dRI_trunc_stats <- data_dRI_trunc %>%
  group_by(., prot) %>%
  summarise(max = max(scaled_norm_intensity))
data_dRI_trunc_stats
```

```
## # A tibble: 4 x 2
##   prot    max
##   <chr> <dbl>
## 1 G363D 0.669
## 2 G401S 0.715
## 3 R710G 0.601
## 4 WT    0.279
```

```r
#compute time of max signal UV
# output = value of max scaled and normalized intensity
# need to parse for time
data_UV_trunc_stats <- data_trunc_UV %>%
  group_by(., prot) %>%
  summarise(max = max(scaled_norm_intensity))
data_UV_trunc_stats
```

```
## # A tibble: 4 x 2
##   prot    max
##   <chr> <dbl>
## 1 G363D 0.842
## 2 G401S 0.894
## 3 R710G 0.716
## 4 WT    0.420
```

```r
#compute time of max signal dRI
# output = value of max scaled and normalized intensity
# need to parse for time
data_LS9_trunc_stats <- data_trunc_LS9 %>%
  group_by(., prot) %>%
  summarise(max = max(scaled_norm_intensity))
data_LS9_trunc_stats
```

```
## # A tibble: 4 x 2
##   prot    max
##   <chr> <dbl>
## 1 G363D 0.606
## 2 G401S 0.680
## 3 R710G 0.875
## 4 WT    0.840
```