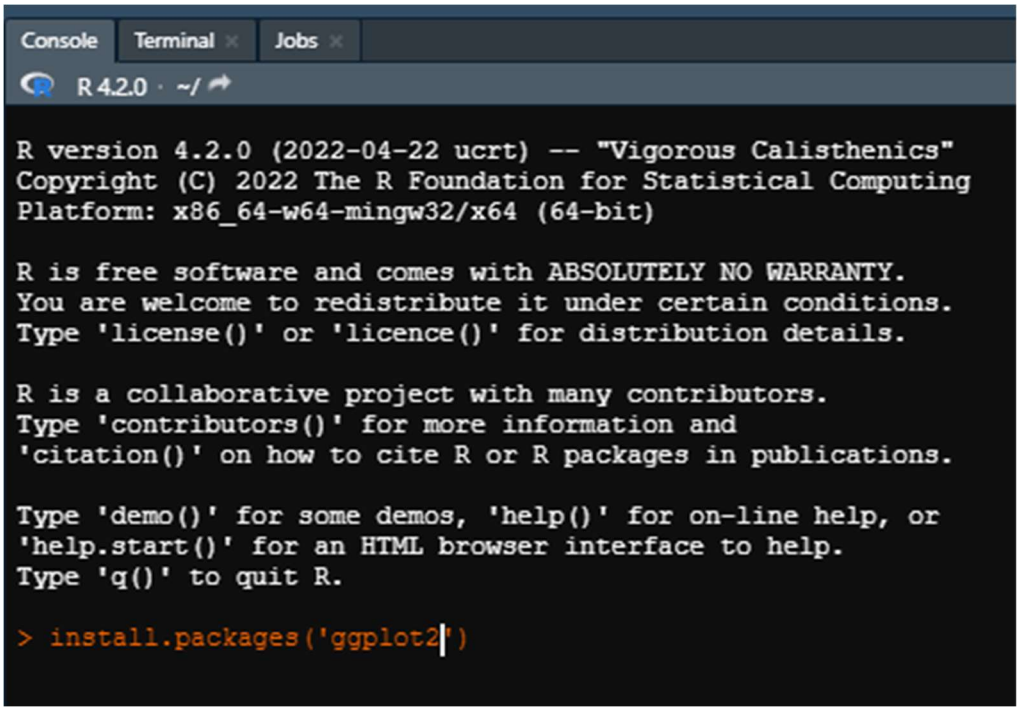
Braden Keiser
PBSA Analysis Protocol:
2024/02/28

0. **Install Rstudio and packages**
   You need to download Rstudio onto your windows computer and install the ggplot2 and stringr packages by issuing into the R terminal at the bottom left of Rstudio:
   --> install.packages('ggplot2')
   --> install.packages('stringr')



   I also recommend installing bio3d for future work with running MD simulation analysis.

1. **Get pbsa.dat file**
   Within the qm_protocol/amber_scripts folder is the getddg_pbsaresults.sh file. Copy this file to your pbsa_t# directory and run it with:

   bash getddg_pbsaresults.sh

   You will now have generated a pbsa.dat file that just displays the various PBSA information. Most importantly, it selects for the delta PBSA value that will be useful for running the Rscript hmg_decomp_pbsa.R.

2. **Download data**

   Download the pbsa.dat and total_ddg.dat

   tar cfvz pbsa_results.tar.gz total_ddg.dat pbsa.dat

   and untar in your mobaxterm:

tar xfvz pbsa_results.tar.gz

3. **Open Rstudio and open hmg_decomp_pbsa.R**

At the args variable, change the three items in the list to:

I. The directory where the data is (replace any '\' with '/')
II. The name of the variant (i.e., WT or mutant name)
III. Open the pbsa.dat and locate DELTA TOTAL at the bottom:

```
Differences (Complex - Receptor - Ligand):
Energy Component            Average              Std. Dev.   Std. Err. of Mean
------------------------------------------------------------------------------
VDWAALS                    -84.5096               4.6764              0.2952
EEL                       -101.5008               8.2635              0.5216
EPB                        160.8258               8.5333              0.5386
ENPOLAR                    -64.2779               1.3742              0.0867
EDISPER                    112.3130               1.9234              0.1214

DELTA G gas               -186.0105               8.1493              0.5144
DELTA G solv               208.8608               8.9531              0.5651

DELTA TOTAL                 22.8503               8.1019              0.5114
```

It is the 22.8503, here. This was a short simulation and hopefully your results are more energetically favorable. Use this value for the third item.

Alternatively, you can run in mobaxterm:

grep "DELTA TOTAL" pbsa.dat

```
28/02/2024 09:34.47 /drives/g/hmg/tested_wt/pbsa_results grep "DELTA TOTAL" pbsa.dat
DELTA TOTAL              22.8503              8.1019              0.5114
```

Therefore, you have in Rstudio:

```
print('Args should be: PWD, MUT, PBSA_DDG')
args=c('g:/HMG/tested_wt/pbsa_results', 'WT','22.8503')
print(paste0('Arguments: ', args))
```

4. **Piecemeal run the script**

Select the first 11 lines up through the dg <- read.table:

```
1  library('ggplot2', 'stringr')
2  rm(list=ls())
3  args <- commandArgs(trailingOnly=TRUE)
4  print('Args should be: PWD, MUT, PBSA_DDG')
5  args=c('g:/HMG/tested_wt/pbsa_results', 'WT','22.8503')
6  print(paste0('Arguments: ', args))
7  setwd(args[1])
8  total_ddg <- args[3]
9
10 dg <- read.table('total_ddg.dat', header = TRUE, sep = ',',
11                  as.is = TRUE) #MUST DO AS IS OTHERWISE IT SORTS RESNAMES ALPHABETICALLY
```

Run this section by pressing:

Alt + Enter

If you get this:

```
Error in read.table("total_ddg.dat", header = TRUE, sep = ",", as.is = TRUE) :
  more columns than column names
```

Open the total_ddg.dat file in Kate and **remove** all lines BEFORE (**ABOVE** the red line):

```
10  ALA
    727,15.44908366533
    .924975502703476,0.
    71,7.680282829896699
11  Residue,Location,Int
12  #,,Avg.,Std. Dev.,St
    Mean,Avg.,Std. Dev.,
13  PHE   2,R PHE
```

This “Residue…..“ starts our decomposition column names and is followed by all residues that were found to be within 6 angstroms of the HMG ligand in pt5 running the nearlig.py script.

We also need to **remove** all lines AFTER the last residue in this list (**BELOW** the red line):

```
HMG 786,L HMG
1,0.0,0.0,0.0,-42.254828685258964,2.3382380311
960159366,3.805583345511449,0.240206290820904,
5358,0.1867739468159517

Sidechain Energy Decomposition:
```

Additionally, **remove any blank lines** above and below the red lines. Save the document, and we can run all lines in the Rscript.

CTRL + A to select all lines in the Rscript in Rstudio, and then hit Alt + Enter to run everything.

After running the script, you should see a png file in the folder. The plot should also open in Rstudio. There are many residues that are associated with this ligand, as it's very large. You can change the plotting parameters for the cutoff DDG values by altering the numbers in the following line:

```
    res_list <- c(new_line)
  }
if (dg_df$TOTAL[i] < 0.05 & dg_df$TOTAL[i] > -0.05) {
    print('data too inconsequential; will not be added to graph')
} else {
    if (exists("fin") == TRUE & exists("new_reslist")==TRUE) {
        print('data sufficiently large, adding to graph')
        new_row <- c(toString(dg_df$res[i]), dg_df$TOTAL[i])
        fin <- rbind(fin, new_row)
        new_reslist<-append(new_reslist,new_line)
    } else if (exists("fin") == FALSE & exists("new_reslist")==FALSE) {
        print('making fin dataframe')
        fin <- data.frame(fin_res = dg_df$res[i], fin_tot = dg_df$TOTAL[
        new_reslist <- c(new_line)
    }
```

This just says to not plot residues that have energies between -0.05 – +0.05. These are relatively insignificant. Perhaps, others may not be necessary. You could also manually remove certain residues if they are not of interest to you.

The residues selected for decomposition do not contribute to or affect the value obtained during the PBSA calculation.