vs_sup

drsimonturega

2025-12-16

Virtual screening for high affinity guests for synthetic supramolecular receptors

Orginal literature

https://doi.org/10.1039/C5SC00534E

Function used in this analysis

I don't have a R data science package(library) yet...

Replication of our analysis in the manuscript using python

We load and clean data from the supporting information

```
df = read.csv("tab_gold_wt.csv", header = TRUE)
```

Run some exploratory data analysis

Data summary

```
summary(df)
```

```
##      Guest     Ligand_clash     Ligand_torsion    Part_buried
##  Min.   : 1   Min.   :0.00000   Min.   :0.0000   Min.   :-4.621
##  1st Qu.:18   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:-3.046
##  Median :35   Median :0.00000   Median :0.0000   Median :-2.193
##  Mean   :35   Mean   :0.02582   Mean   :0.0650   Mean   :-2.027
##  3rd Qu.:52   3rd Qu.:0.00000   3rd Qu.:0.0016   3rd Qu.:-1.708
##  Max.   :69   Max.   :1.78130   Max.   :1.0855   Max.   : 8.908
##    Non.polar      Ligand_flexibility    logKexp
##  Min.   :-72.20   Min.   :0.0000     Min.   :-1.000
##  1st Qu.:-49.49   1st Qu.:0.0000     1st Qu.: 1.860
##  Median :-40.64   Median :0.0000     Median : 3.600
##  Mean   :-42.43   Mean   :0.7971     Mean   : 3.167
##  3rd Qu.:-34.32   3rd Qu.:1.0000     3rd Qu.: 4.300
##  Max.   :-22.77   Max.   :7.0000     Max.   : 8.000
```

we still have guest number the data frame we will remove it

```
df <- df[-c(1)]
summary(df)
```

```
##   Ligand_clash     Ligand_torsion    Part_buried       Non.polar
##  Min.   :0.00000   Min.   :0.0000   Min.   :-4.621   Min.   :-72.20
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:-3.046   1st Qu.:-49.49
##  Median :0.00000   Median :0.0000   Median :-2.193   Median :-40.64
##  Mean   :0.02582   Mean   :0.0650   Mean   :-2.027   Mean   :-42.43
##  3rd Qu.:0.00000   3rd Qu.:0.0016   3rd Qu.:-1.708   3rd Qu.:-34.32
##  Max.   :1.78130   Max.   :1.0855   Max.   : 8.908   Max.   :-22.77
##  Ligand_flexibility    logKexp
##  Min.   :0.0000     Min.   :-1.000
##  1st Qu.:0.0000     1st Qu.: 1.860
##  Median :0.0000     Median : 3.600
##  Mean   :0.7971     Mean   : 3.167
##  3rd Qu.:1.0000     3rd Qu.: 4.300
##  Max.   :7.0000     Max.   : 8.000
```

```
col_names = colnames(df)
print(ncol(df))
```

```
## [1] 6
```

```
print(col_names[1])
```

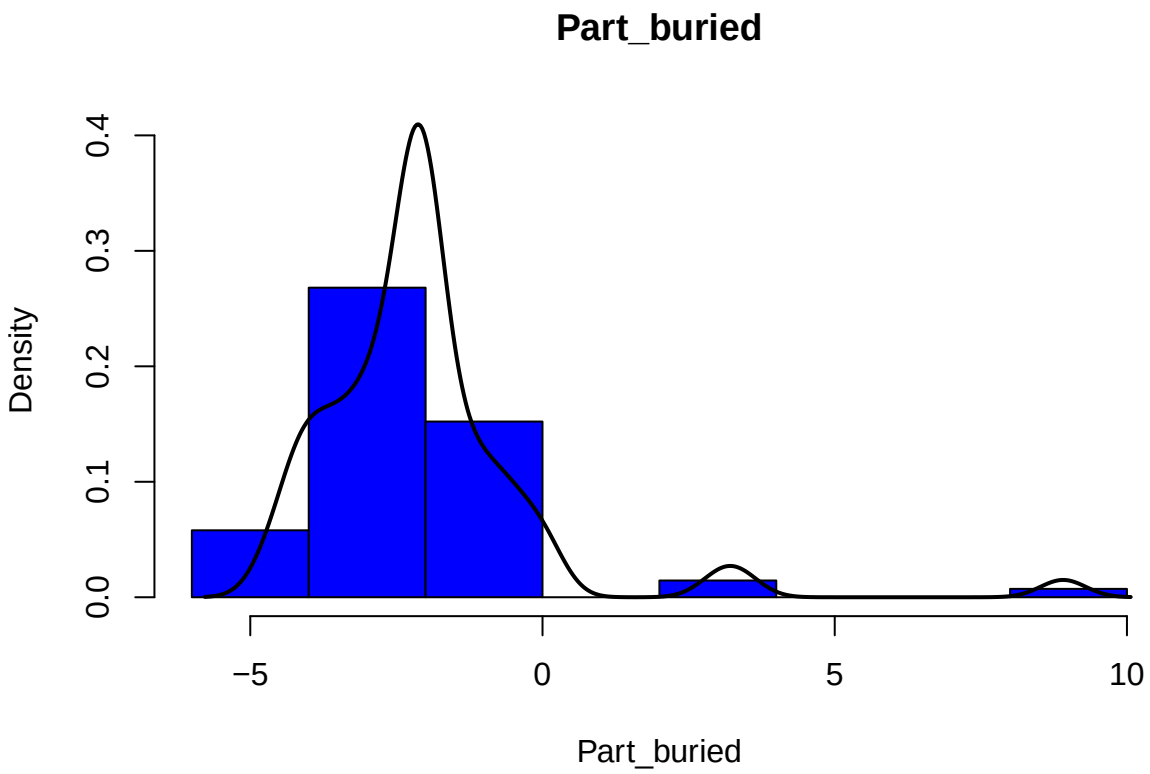
```
## [1] "Ligand_clash"
```

```
int_made_even(3)
```

```
## [1] 4
```

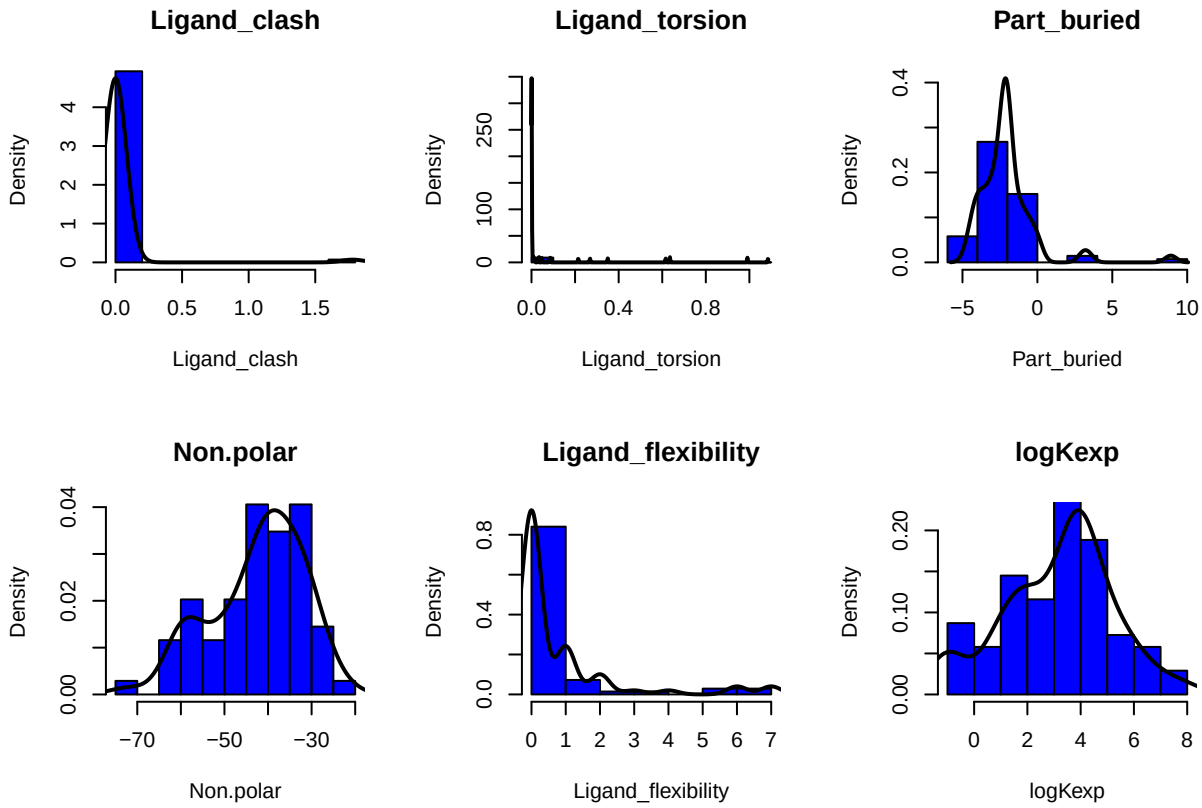
Colored Histogram with KDE line for clarity

```
hist_kde(df, "Part_buried")
```




```
multi_plot(df, hist_kde)
```

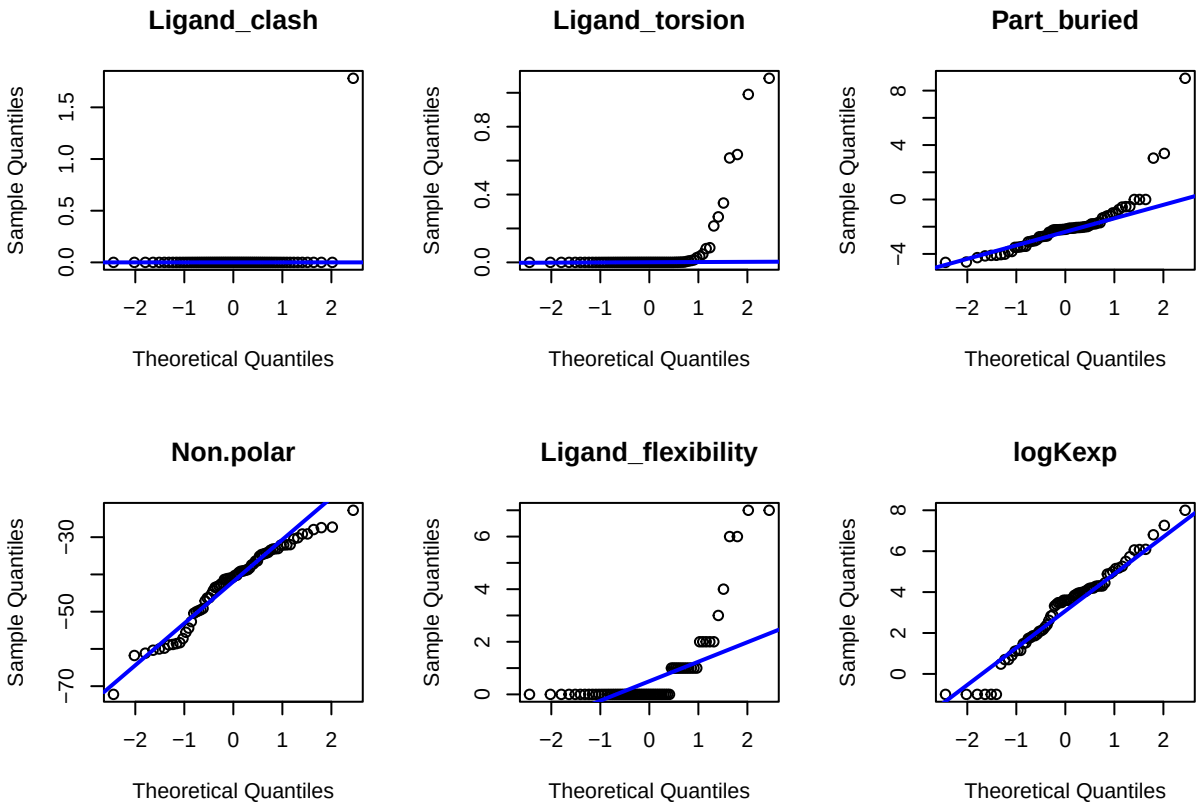
```
## [1] 6
## [1] 6
## [1] 6
```



Quantile quantile plots Our quantile quantile plots show us weather our data fits to a theoretical distribution

```
multi_plot(df, q_q_plot)
```

```
## [1] 6
## [1] 6
## [1] 6
```



This is experimental data from 69 individual data points our team generated so we don't really expect it to fit a normal distribution. We would expect the replicates of **logKexp** to fit a normal distribution and the error quoted in the manuscript is at 95% confidence.