Q1)

Reported DNA, RNA, and Hybrids sequences and their size distribution.

1. Reported by Date, stack by DNA and RNA

Q2)

Sequences with any reported binding and break them into RNA and DNA to report trends in binding by doing a statistical analysis.

Q3)

If any proteins bind to the same target, how does their binding affinity compare?

Approaching the data analysis

Steps

Q1) Create a data frame from the data.

DF = ID, Resolution, Date, Size, DNA, RNA, and Hybrid

Q2) Drop any rows that don’t have a value in ligand and value column. From here, we will get either DNA or RNA sequences with KD values. We will do statistics to look for trends in DNA vs RNA sequences.

3) I need to clean up and present csv with KD values and we will be able to merge on the original database with ligand ID. Compare protein vs DNA and RNA binding. Multimodal binding of proteins vs DNA and RNA.