Instructions for running the code on your computer:

This block of code must be changed so the file directories reflect the actual location of the folders named “On Target Output”, “Off Target Output”, and “Runs”. This will involve creating a folder named “Runs” where the final product will be output.

on\_base\_dir = 'D:\\DJ\_Files\\On Target Output\\'  
off\_base\_dir = 'D:\\DJ\_Files\\Off Target Output\\'  
path = 'D:\\DJ\_Files\\Runs\\'

Other Information:

The folder named “Other Stuff” is not necessary to run the final product. It contains multiple rough drafts and test files for the final product. It also contains several programs dedicated to creating the data sets that are used in the final program.

Within the “On Target Output” and “Off Target Output” folders there are many files. These files are named according to the convention Group\_Family\_Name\_Drug\_PDB ID. This allows a user to easily find kinases of specific names, families, or groups. The Excel spreadsheets in the folders are named the same and contain an average Interaction Fingerprint(IFP), structure data listing the number of bound configurations, bioactivity data, structure coordinates for each configuration, and IFPs for each configuration.

Within the folders “On Target Composites” and “Off Target Composites” there are Excel spreadsheets named for different kinases. These spreadsheets have a composite for all drugs that bind to that kinase, followed by one sheet for each drug that binds to that kinase. These sheets contain graphs at the bottom to allow for easy interpretation of the data.