* ‘exp\_affinities\_ds.csv’:

A dataset containing pairs of proteins along with their experimental affinities as mined from PDBbind v2020118, SKEMPI 2.0119, MPAD120 and Binding Affinity Benchmark version 2 (BABv2). Affinities are provided in the form of (kJ/mol) in the column ‘DG (kJ/mol)’

* ‘dg\_processed\_dataset.csv’

Based on the experimental affinities from ‘exp\_affinities\_ds.csv’, 61 features were calculated for the regression analysis. The column ‘DG’ contains the experimental affinities normalized and converted to 0-1 range.

* ‘dg\_test\_ds.csv’, ‘dg\_train\_ds.csv’

The test and train datasets used in the regression analysis. An an 80-20 training-test ratio was implemented, ending with 415 samples in the training set and 104 samples in the testing set.