## Single Variation based on Co-Variation data

1. Re-write Data Files (ReWriteDataFiles)
   1. Unpacks the Data files
   2. Builds many small files in one folder per experiment, e.g. „1\_455.txt“ contains all covariation data for experiment 1, position 455
2. Clear Lines (ClearLines)
   1. Clears empty lines in the data files. These are the lines where pos 1 <= pos 2
3. Estimate Error (CoVarErrorAnalysis)
   1. Generates a file with the error estimates “ErrorEstimates.mat”
4. Compute raw KD values (SingleVarAnalysis\_BasedOnCoVarDataXX\_XX\_ComputeKDs)
   1. Generates a file named “Results.mat” that contains the raw KD values, the position weights (% of maximum coverage), signal to noise ratio, and the number of reads for each combination of co-varying positions.
5. Plot KD values and apply quality criteria (SingleVarAnalysis\_BasedOnCoVarData11\_02\_PlotKDs)
   1. Generates all plots and writes the output into a ‘PositionWiseKDEstimates.csv’ file
   2. Maximally affecting mutations: ‘PositionWiseKDEstimatesSummary.csv’, ‘IdentifiedPositionsFullSet.csv’
   3. Calls ‘TestPositionWiseDifferences’ and outputs ‘PositionWiseKDDifferences.csv’
6. SingleVar 🡪 Base-pairedness (SingleVarStructuralAnalysis.m)
   1. Compute/assess base-pairedness, outputs ‘Figure4.csv’

## Co-Variation based on Tri-Variation data

1. Re-write Data Files
   1. Unpacks the data files
   2. Builds many smaller files in one folder per experiment, e.g. „200\_255.txt“ that contains all tri-variation data for experiment 1, between position i = 200, j = 255 and z = \*.
2. Compute for each potentially interacting pair the raw KD values (CoVarAnalysisScript)
   1. Extract important residues (where Kd\_m > 1 for any m). 🡪 ‘ImportantResidues.csv’.
   2. Assess potential Partners (Important residues that could form a WC, or WB pair), for each important residue

🡪 ./ResultsTriVatWCWBnew/XX\_partners.csv , where XX is the position of interest.

Output the number of Pairs

🡪 ./ResultsTriVatWCWBnew/NumberPairs.csv

* 1. Compute for each position i and its potential partner j raw Kd values Kd\_m\_w\_w, Kd\_w\_m\_w and Kd\_m\_m\_w, the position weights (% of maximum coverage), signal to noise ratio, and the number of reads for each combination of co-varying positions.

🡪 ./ResultsTriVatWCWBnew/XXi\_XXj.mat, where Xxi and XXj are the positions of interest i and j

1. Plot Epistasis values, apply quality criteria, evaluate statistically & identify stems (InferEpistasis).
   1. Writes EpistasisEstimates.csv.