

## Correlation of Fitness Effects

The ruggedness of a ribozyme family can be measured using the fitness correlation  $\gamma_d$ , which is the average correlation of activity effects of single mutations in  $d$ -mutant neighbors (where  $d$  is the Levenshtein edit distance, i.e., the number of substitutions, insertions or deletions between two related sequences).

The correlation of fitness effects,  $\gamma_d$ , is calculated following [Ferretti et al., 2016](#).

### How to use the script to calculate correlation of fitness effects:

To reproduce the numerical results reported in the publication, the python script `ActivityCorrelationGamma.py` can be run as:

```
python ActivityCorrelationGamma.py input_file sheet_name correlation_distance
```

where `input_file` corresponds to the excel file `ActivityObservedData.xlsx` and `sheet_name` corresponds to either Family\_2.1, Family\_1A.1, Family\_1B.1, Family\_1B.2 or Family\_1A.2. `Correlation_distance` is an integer number {0..4}.

The excel source file `ActivityObservedData.xlsx` can be found in this repository. Each spreadsheet in the excel file corresponds to a different family and follows the same format: column A) sequences, column B) calculated activity (including values under the baseline activity), column C) calculated activity if above the baseline activity (or baseline activity if below), and column D) logarithm of the values in column C).