**Date:** October 10, 2025  
**Language:** Python 3.10.6  
**Requirement:** Python 3+

**Overview:** The theory, rationale, and specifics of data analyses utilized in this calculator are explained in more detail in the associated 2018 publication.

Hodges, A. M., A. W. Fenton, L. L. Dougherty, A. C. Overholt, and L. Swint-Kruse. 2018. 'RheoScale: A tool to aggregate and quantify experimentally determined substitution outcomes for multiple variants at individual protein positions', Hum Mutat, 39: 1814-26.

Additions to the 2025 revised calculator are described below in red.

RheoScale 2.0 is a Python-based analysis tool that classifies protein positions based on quantitative variant data. It reads a CSV file containing measured values (e.g., enzyme activity, fluorescence, binding, etc.) for each amino acid substitution and assigns each position to classes such as: Neutral, Rheostat, Toggle, Moderate, Adverse, Enhancing, WT/inactive. The script also generates histograms and a summary output file that can be used for further analysis.

**Position classes:**

At least 5 amino acid variants are required for each position to be analyzed. RheoScale first uses the value and error for the wild-type variant to determine whether a position is “neutral” or “non-neutral”. The variant sets for non-neutral positions are then assessed with a modified histogram analysis. The logic used to define position classes follows the order: Neutral>Toggle>Rheostat>Moderate/Adverse/WT-Inactive\_split/Enhancing. The descriptions of these position classes, as well as the publications describing their default score thresholds, are:

**>>Neutral**: At least 70% of substitutions have WT-like outcomes for the parameter measured. The neutral score for each position, which is used to assign neutral positions, is calculated separately from all other scores, by using a neutral bin that is centered on the wild-type value and is (usually) independent of the histogram bin size.

Reference: Martin, Tyler A., Tiffany Wu, Qingling Tang, Larissa L. Dougherty, Daniel J Parente, Liskin Swint-Kruse, and Aron W. Fenton. 2020. 'Identification of biochemically neutral positions in liver pyruvate kinase', *Proteins: Structure, Function, and Bioinformatics, 88: 1340-50*

**>>Moderate rheostat**: Substitutions have non-neutral outcomes, but the set of substitutions samples less than half of the possible range AND the values are closer to WT than to the "dead" end of the range.

Reference: Swint-Kruse, L., T. A. Martin, B. M. Page, T. Wu, P. M. Gerhart, L. L. Dougherty, Q. Tang, D. J. Parente, B. R. Mosier, L. E. Bantis, and A. W. Fenton. 2021. 'Rheostat functional outcomes occur when substitutions are introduced at nonconserved positions that diverge with speciation', *Protein Sci, 30: 1833-53.*

**>>Rheostat**: Simplistically, the position's set of substitutions samples at least half of the possible functional range. When the "weighted" rheostat score is used, the sampling might be a little less than half of the range, but contributions from variants with partial loss-of-function are weighted more heavily, since they provide more confidence that a position is a rheostat position. The weighted score has been used in almost all studies to date. Rheostat behavior cannot be detected from the average of a position's substitution values.

Reference: Hodges, A. M., A. W. Fenton, L. L. Dougherty, A. C. Overholt, and L. Swint-Kruse. 2018. 'RheoScale: A tool to aggregate and quantify experimentally determined substitution outcomes for multiple variants at individual protein positions', *Hum Mutat, 39: 1814-26.*

**>>Adverse**: Like Moderate rheostat positions, substitutions have non-nuetral outcomes and the set of substitutions samples less than half of the possible range. However, the set of values are closer to the "dead" end of the range than to WT.

Reference: Sreenivasan, S., Fontes, J. D., and Swint-Kruse, L. 2025. 'Dissecting the effects of single amino acid substitutions in SARS-CoV-2 Mpro', Protein Science 34, e70225.

**>>Toggle**: Around 2/3 of the substitutions lack detectable activity.

Reference: Miller, M., Y. Bromberg, and L. Swint-Kruse. 2017. 'Computational predictors fail to identify amino acid substitution effects at rheostat positions', *Sci Rep, 7: 41329.*

**>>WT/Inactive split**: Half of the substitutions have WT-like outcomes and the other half lack detectable activity. This may be a hallmark of altered protein stability. These are a special case of "binary" positions (for which substitutions fall in only 2 bins).

Reference: Page, B. M., T. A. Martin, C. L. Wright, L. A. Fenton, M. T. Villar, Q. Tang, A. Artigues, A. Lamb, A. W. Fenton, and L. Swint-Kruse. 2022. 'Odd one out? Functional tuning of Zymomonas mobilis pyruvate kinase is narrower than its allosteric, human counterpart', *Protein Sci, 31: e4336.*

**>>Enhancing**: At least 80% of substitutions enhance the measured parameter relative to the upper limit of the neutral bin.

Reference: Sreenivasan, S., Fontes, J. D., and Swint-Kruse, L. 2025. 'Dissecting the effects of single amino acid substitutions in SARS-CoV-2 Mpro', Protein Science 34, e70225.

**Overview of Histogram Analyses:**

a) For linear-scale data that spans 1 order of magnitude, histogram bins should be calculated using non-transformed (linear) calculations. If a data set covers more than two orders of magnitude, it should be converted to a log scale. Any functional value that has already been converted to a log scale (e.g., Gibbs free energy) or doesn't cover more than two orders of magnitude (e.g., Hill number) should not be converted. A data set that spans between 1 and 2 orders of magnitude should be carefully considered to determine whether a log scale is appropriate or not. Note that most high-throughput data (e.g., "deep mutational scanning") are already reported in log scale, and should not be converted.

b) The value that represents a completely nonfunctional protein ("dead") must be defined as either the minimum or maximum value used to analyze the data set.

c) The default width of the neutral bin is set to 4 times the WT error (or 4 times the error override), to represent the WT value +/- 2 standard deviations. If no error is designated, the default is set to 2 times the rheostat/toggle bin size. See Martin et al, 2020, for the statistical reasoning. The width of the neutral bin can also be set by the user.

c) The histogram range is the most important parameter for assigning position phenotypes. The range is determined from the minimum and maximum values of the dataset and/or known for the assay. The min and max values of each dataset are calculated automatically. If a data set is known to have min or max values that differ from the experimental dataset, override values can be entered. Various examples that require override values are as follows:

* For datasets with variants that are "better" than wild-type, it can be useful to set the relevant max/min override so that the last bin is populated by at least 5% of the total variants; this prevents a tiny set of highly-active variants from making the range artificially large and thus dominating RheoScale assignments.
* The max/min override cells are used when the dataset being analyzed spans a smaller range of functional values than is known to be possible (e.g., when the dataset does not contain a “dead” variant.)
* The max/min overrides may also be used when investigators wish to designate a “dead” threshold that falls inside the measurable assay range (e.g., any variant with less than 10% activity should be considered “dead”).
* Examples for estimating "dead" values are described in the following publications (as well as others from the Swint-Kruse lab): 1. Hodges, Fenton, Dougherty, Overholt, and Swint-Kruse. 2018. 'RheoScale: A tool to aggregate and quantify experimentally determined substitution outcomes for multiple variants at individual protein positions', Hum Mutat, 39: 1814-26. 2. Sreenivasan, Fontes, and Swint-Kruse. 2025. 'Dissecting the effects of single amino acid substitutions in SARS-CoV-2 Mpro', Protein Science 34, e70225. 3. O'Neil, Swint-Kruse, and Fenton (2024) Rheostatic contributions to protein stability can obscure a position's functional role, Protein Science 33, e5075.
* If any data point falls outside of the max or min override value, it will be reassigned to be the override value for calculations.

d) An error override option is provided for data sets that do not have an error value associated with each functional value. Alternatively, one error value may better represent the error inherent in the experimental methodology. This value is used in determining a recommended number of bins for analyzing the data set. If error override is included for a data set that is converted to a log scale within the calculator, then the error value entered will be propagated using the wild type as the reference value. The formula for error propagation for log calculations in this case is 0.434\*error/[WT value]. If an alternate approach to error propagation is desired (percent error, etc) then the user should amend the data set before including in the calculator.

e) The recommended number of bins is determined through a combination of the average error for the data set as well as the total number of variants at each position. A perfect rheostat position would occupy 20 bins, but error and the number of variants available for each position constrain the number of bins that should be used. The algorithm for the bin number recommendation c is explained in further detail in Hodges et al., 2018. If a different number of bins is desired, that number can be entered with an override. Empirically, 10 bins appear to work well for many datasets. Iterations with different bin numbers show that many position assignments are not very sensitive to the bin number.

**How to use this python version of the RheoScale calculator**

**Python Requirements:** You need **Python 3 or later** and a few standard Python packages.If you don’t have Python installed:

* Go to https://www.python.org/downloads/
* Download the latest **Python 3** version (or higher).
* During installation, **check the box** that says **“Add Python to PATH”**.

To confirm it’s installed, open a terminal (Command Prompt on Windows, Terminal on macOS/Linux), and run:

python --version

You should see something like Python 3.10.6.

**Installing required packages:** Open a terminal (Command Prompt on Windows, Terminal on macOS/Linux), and run:

pip install pandas numpy matplotlib

That’s all the setup you need.

**Input file format:** Your data must be a **CSV file** (comma-separated values) with the following columns:

|  |  |  |  |
| --- | --- | --- | --- |
| **Position** | **Substitution** | **Value** | **Error** |
| 45 | A45V | 0.75 | 0.05 |
| 45 | T | 1.10 | 0.07 |
| ... | ... | ... | ... |
| WT | WT | 1.00 | 0.03 |

* **Position:** The residue number (e.g., 45).
* **Substitution:** Mutation identifier (e.g., A45V or just V). Unlike the Excel version of the calculator, this column must be filled.
* **Value:** Measured value for that variant.
* **Error:** Associated error value. This column can be blank if the error is instead estimated for the whole dataset.
* A value for the reference protein **must** be present in the input data file. It can be any place within the dataset, and it must be labeled “WT” or “wild-type”.

**Running the script:** Place the input file (for example data.csv) and RheoScale script (RheoScale\_2.0.py) in the same folder, then run:

python RheoScale\_2.0.py data.csv  
  
The script will read your file, perform all calculations, and save:

* A folder containing **histogram plots**
* A CSV file containing **RheoScale output results**
* A JPG file distribution of different position classes **in a pie chart**

**Optional flags** are used to customize input and analysis parameters.

Example:

python RheoScale\_2.0.py data.csv --log\_scale False --toggle\_score\_threshold\_override 0.6 --plot\_prefix Mpro

**Flag Options**

|  |  |  |  |
| --- | --- | --- | --- |
| **Option** | **Description** | **Default** | **Input Options** |
| --log\_scale | Convert values to log10 for calculations | True | True/False |
| --min\_override | Force a minimum value cutoff | False | True/False |
| --min\_val | Value to use for minimum override | Requires user value when min\_override = True | Any number (float data type) |
| --max\_override | Force a maximum value cutoff | False | True/False |
| --max\_val | Value to use for maximum override | Requires user value when max\_override = True | Any number (float data type) |
| --error\_override | Override all error values | False | True/False |
| --error\_val | Error value to use if overriding | Requires user value when error\_override = True | Any number (float data type) |
| --bin\_override | Manually set number of histogram bins | Requires user value | Integer between 0-20 |
| --dead\_value | Defines whether the value corresponding to total loss of signal (“dead”) is “Min” or “Max” | Min | Min/Max |
| --neutral\_binsize\_override | Manually set neutral bin size | Auto (2× default bin size) | Any number (float data type) |
| --enhancing\_score\_threshold\_override | Threshold for classifying Enhancing positions | 0.8 | Any number (float data type) between 0 and 1 |
| --neutral\_score\_threshold\_override | Threshold for classifying Neutral positions | 0.7 | Any number (float data type) between 0 and 1 |
| --rheostat\_score\_threshold\_override | Threshold for classifying Rheostat positions | 0.5 | Any number (float data type) between 0 and 1 |
| --toggle\_score\_threshold\_override | Threshold for classifying Toggle positions | 0.64 | Any number (float data type) between 0 and 1 |
| --plot\_prefix | Prefix for all output filenames | None | Text |

**Output files:**

After running, you’ll find two output files:

1. A folder called Plots/ or <prefix>\_Plots/ contains histograms in .jpg format for each position and a plot of “All” variants.
2. A CSV file: output-data.csv or <prefix>\_RheoScale-output.csv

This files lists the analysis parameters used and then lists each position and its computed parameters, in the format:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Position** | **Variants** | **Enhancing** | **Neutral** | **Unweighted Rheostat** | **Weighted\_Rheostat** | **Toggle** | **Average** | **Std\_dev** | **Assignment** |
| 45 | 10 | 0.12 | 0.80 | 0.60 | 0.65 | 0.10 | 0.22 | 0.91 | Neutral |
| 46 | 8 | 0.00 | 0.29 | 0.60 | 0.65 | 0.15 | 1.47 | 0.77 | Rheostat |
| ... | ... | ... | ... |  | ... | ... | ... | … | … |

**Troubleshooting tips:**

|  |  |  |
| --- | --- | --- |
| **Problem** | **Likely Cause** | **Solution** |
| “Error: NO WILD TYPE GIVEN IN DATA” | WT row missing in input data | Add a WT entry with its value and error |
| “Error: The file was not found” | Wrong filename or path | Make sure the CSV file is in the same folder or use full path |
| Plots not generated | Matplotlib not installed | Run pip install matplotlib |
| Output CSV is empty | Fewer than 5 variants at positions | Add more variants or check input data |