QSPR.pxp is a packed WaveMetrics Igor Pro 8 experiment file containing the macros and compiled functions to perform regularized (multivariate) multiple linear regression ((M)MLR) and univariate logistic regression (LGR) on the data that require molecular descriptors as explanatory variable, in addition to other numerical parameters. The descriptors can be introduced from an HTF5 file that is generated by a program *chemical\_descriptor.py* in

<https://github.com/AD-SDL/redoxmer_python_kinetics_analysis>

An example of this input is given in *des.h5* file. There are also two \*.csv formatted input examples for MLR and LGR regressions.

The program requires *HDF5-64.xop* extension to read HDF5 descriptor files *(\*.h5)*. The first time it is used, it will also ask to indicate the folder where procedure *(\*.ipf)* files are located.

Here is what these procedure files do:

*file\_io.ipf and service.ipf* procedures contain service functions

*gen\_discrete.ipf* contains the Genetic algorithm routines

*qspr.ipf* is the main procedure file containing MMLR and MLR regression

*qspr\_cross\_validation.ipf* contains the programs for cross validation of the global solutions using random splitting of the data

*LinReg.ipf* contains Igor implementation of linear algebra for regularized linear and logistic regression

The macros and procedures are available from **Macros** and **QSPR** drop menus at the top of the work window. The most important of these are MLR\_Genetics and LGR\_Genetics macros in Macros.

*QSPR->Descriptors->CreateSet\_HDF5\_des* will read the des.h5 file with molecular descriptors

*QSPR->CVS import->LoadCSV*\_table will read input .csv file with tabulated values to explain, and all additional explanatory variables (parameters).

A user sets the target number of descriptors. The program performs genetic algorithm optimizations seeking to find a set of descriptors of this size to minimize the rms or its logistics analog globally. Hashing and starring is used to exclude descriptors or make their use obligatory (*QSPR>Hash and Star*). The program has multiple visualization options, including making plots, tables, and MP4 formatted videos. *QSPR->Sets* is a suite of macros to save, export and import sets, consisting of input data and fits.

The questions on how to use this software can be addressed to Eli Shkrob

[shkrob@anl.gov](mailto:shkrob@anl.gov)

[shkrobius@gmail.com](mailto:shkrobius@gmail.com)