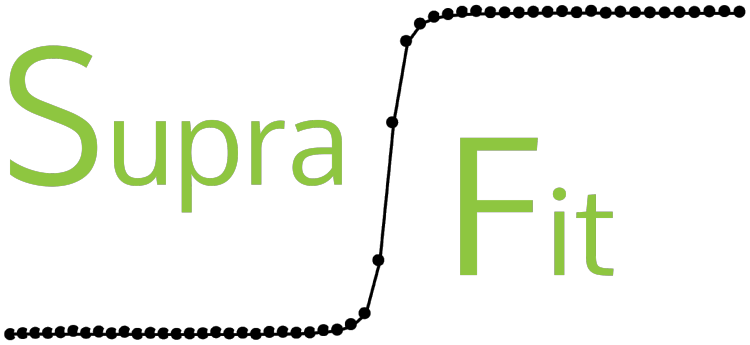


SupraFit Quickstart

Version 2 Beta 2 (1.8)

Conrad Hübler

August 26, 2019

Preface

Supramolecular titration experiments are widely used to determine receptor-substrate stability constants. Despite the popularity of these experiments, sometimes older stand-alone software is used, for example WinEQNMR[1] and HypNMR[2] for NMR titration. Recently alternative software, based on Matlab,©[3] has been made available. Additionally, a web service[1] can be used for analysis.[4, 5] A commercial Matlab and Excel based solution called ReactLab™ is offered by Jplus.[2] In case of Isothermal Titration Calorimetry (ITC), the above mentioned software can not be used - however other programs like NanoAnalyze[6] or pytc[7] are available. Different programs focusing on different aspects of supramolecular titration experiments provide statistical approaches, making it even more difficult to compare the results with other experiments.

To have a modern offline application with consistent statistical approaches that is user-friendly yet powerful and fully free, SupraFit came to my mind. The development of SupraFit started with the Tutorial Review by Thodarson[8] and the provided Matlab scripts. The whole software is based on C++ and the modern graphics toolkit Qt.[9] The mandatory mathematical part, the nonlinear fitting, is done with the open source library Eigen.[10] Therefore, apart from Qt and Eigen, no third party libraries are needed to compile and run SupraFit on Linux and Windows systems (Mac OS has yet not been tested).

A brief overview on SupraFit features:

- Support for NMR and UV/VIS Titration Experiments
  - 1:1, 2:1/1:1, 1:1/1:2 and 2:1/1:1/1:2 titration models including BC_{50}^{0} values[11]
  - Full control over cooperative binding models[12]
  - Local and global fitting
- Support for Isothermal Titration Calorimetry
  - NMR Titration analogous ITC models
  - Additional dilution correction via linear regression
- Support for Michaelis-Menten Kinetics and Monomolecular Kinetics
- Combining different experiments to *Meta Models*
- Visual comparison of different fitted models on the same dataset

[1] http://supramolecular.org/ as part of the OpenDataFit Project.
[2] http://jplusconsulting.com/

- Statistics and further analysis
  - Monte Carlo Simulation[13]
  - Resampling methods
  - Error based analysis[13, 14]
  - Akaike's Information Criterion ("Entropy" based)[15, 16]
- Simple global optimisation algorithm
- A modern (colourful) and easy-to-use user interface
- Easy loading, saving and exchanging of projects
- Model simulation including gaussian random error
- PNG export of charts
- Multithreading support

Acknowledgements

Special thanks to Prof. Monika Mazik for her support to develop this application, the GraFa Freiberg for the financial support. Thanks to Stefan Kaiser for being the first person, who uses this program (being the α and β tester). Thanks to Sebastian Förster for ideas and feedback.

Thanks to Alexander König for proofreading this Quickstart and helpful comments.

Licence [17]

Contents

1 Quick Start

Prior to all

The quickstart focuses on the main user interface features, that one can easily use SupraFit, without reading about theoretical background or dealing with too many possible options and tweaks. Many buttons have tooltips, which support the user. They can be enabled and disabled in the *Configure* dialog.

1.1 Download and installation

Binaries

Windows binaries are provided for the current development version on the github repositorium. They are linked against Qt 5.13, which will be provided as dynamic linked libaries (*.dll) as well. Binaries for linux will not be provided, but can be generated easily from the source code.

Compile from source

SupraFit can be obtained free of charge from Github by cloning the master branch (git is being expected)

```
git clone --recursive git@github.com:conradhuebler/SupraFit.git
```

To compile SupraFit, a C++11 capable compiler, CMake 3.0 or newer and at least Qt 5.10 are mandatory. SupraFit itself has been tested with Qt 5.10, Qt 5.12 and Qt 5.13, compiled with GCC 5, GCC 6, GCC 7, GCC 8 and Clang on linux platforms as well as with MinGW 5 and Microsoft Visual Studio on Windows systems.

```
cd suprafit
mkdir build
cd build
cmake .. -DCMAKE_BUILD_TYPE=Release
make
```

The binary is then located in the build directory. To start SupraFit just type

```
./suprafit
```

in the build directory.

A command line tool for simulation is available as well using the command

```
./suprafit_cli
```

The downloaded source code from the release page on github doesn't provide fisher_dist, libpeakpick and Eigen, which are expected to be in the *external* directory.

```
...
 external
        /fisher_dist
        /libpeakpick
                /eigen
                /libpeakpick
 src
 ...
```

fisher_dist can be downloaded from the fisher_dist github repositorium[3] and libpeakpick from the the libpeakpick github repositorium.[4] Both, SupraFit and Libpeakpick require Eigen to be located in the LibPeakPick directory. Eigen can be downloaded from the Eigen Website,[5]

1.2 Importing

Nomenclature

Supramolecular chemistry is all about hosts, guests, substrats and receptors. These entities are somehow used in an inconsistent way across science. Therefore, for SupraFit the following terms will be defined and used. The substance that is added to another substance is defined as **guest** or **substrate** molecule (B), leaving the other to be the **host** or **receptor** molecule (A).

[3] https://github.com/conradhuebler/fisher_dist
[4] https://github.com/conradhuebler/libpeakpick
[5] http://eigen.tuxfamily.org/index.php

NMR, UV/VIS and Fluorescence Titrations

Although there is technically no limitation on signals or wavelengths that can be processed simultaneously, SupraFit doesn't provide methods for dimension reduction like PCA. Therefore analysis of titrations from UV/VIS or fluorescence should not contain all wavelengths. Prior selection is recommended.

Any data table should consist of at least three columns with the first column being the *host concentration* (A), the second on the being *guest concentration* (B) and then followed by either the shifts of the NMR signals or the intensities at selected wavelength. The first row should hold the "spectrum" of the pure host component (see table 1), wherein no guest is added to the solution. In most experiments, the concentration of the host is kept constant during the titration **and** is the source of the signal. The guest molecule's concentration is increases **and** it does not influence the observed signal directly. It is silent! In UV/VIS Titration the guest does not absorb at the selected wavelength, while in NMR titrations, only the chemical shifts of the receptor are being observed. For more information, see P. Thordarson, *Chem. Soc. Rev.* **2011**, *40*, 1305–1323.

Please note that, at the current stage, fluorescence titration hasn't been tested. It is only basically implemented. The completion is planned.

Table 1: Exemplary table for NMR Titration data. The host concentration is kept constant (first row) and the observed signals belong to the host substance. The guest molecule's concentration changes (second row) and does not affect the signal directly.

Host	Guest	Signal 1	Signal 2	Signal 3	Signal 4	Signal 5	Signal 6	Signal 7
$1 \cdot 10^{-3}$	0	6.33	6.03	2.36	2.22	4.35	3.68	3.19
$1 \cdot 10^{-3}$	$1.33 \cdot 10^{-4}$	6.32	6.05	2.34	2.23	4.3	4.41	3.19
$1 \cdot 10^{-3}$	$2.65 \cdot 10^{-4}$	6.31	6.06	2.33	2.23	4.28	4.82	3.19
$1 \cdot 10^{-3}$	$3.98 \cdot 10^{-4}$	6.3	6.07	2.32	2.23	4.25	5.28	3.19
$1 \cdot 10^{-3}$	$5.31 \cdot 10^{-4}$	6.29	6.08	2.31	2.23	4.23	5.69	3.19

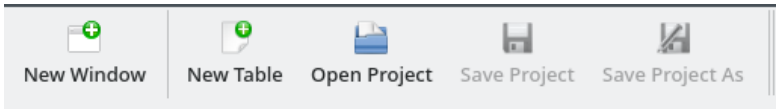


Figure 1: Main import and save actions.

Isothermal Titration Calorimetry

SupraFit handles raw *.itc files, already integrated *.dH files and raw table files:

1. A table of two columns, the first column holding the inject volume and the second one the heat, should be prepared. Concentrations and cell volume can be supplemented after any ITC Model has been added. SupraFit doesn't know of these parameters unless the first ITC Model has been loaded.

2. Loading an already integrated *.dH from Origin is straightforward, stored parameters will be loaded automatically.

3. Loading *.itc and simple *.dat files requires integration, which can be done by SupraFit. Using this approach, a separate dilution experiment can be subtracted automatically. Cell volume and concentrations can be read from the *.itc files!

Michaelis Menten Kinetics

Michaelis Menten Kinetics need the substrate concentration (first column) and the reaction rate (second column).

Monomolecular Kinetics

Monomolecular Kinetics need the time in the first column and the concentration in the second column.

1.3 Loading the table

Open any of these files mentioned above with the **Open Project** button or copy the content of either a whitespace or tabulator separated file into the **New Table** dialog (Fig. 2).

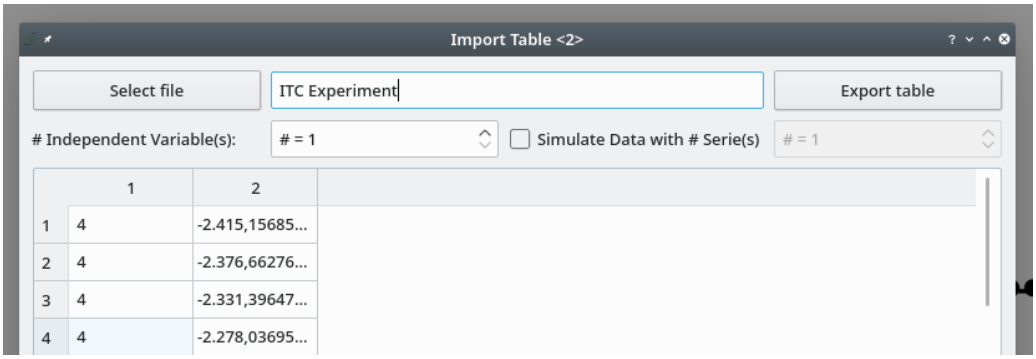
Alternatively, drag files from the file manager into the project list on the left hand side (fig. 3).

Decimal marks such as points or commas are allowed in the tables. They will be treated equally. Set the number of independent variables in the spinbox (fig. 2) to

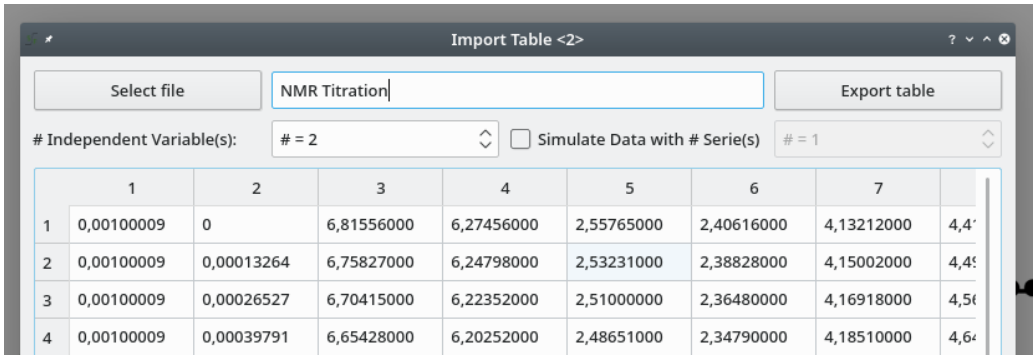
- **1** for Kinetics and ITC models

- **2** for NMR/UV-VIS and fluorescence titrations

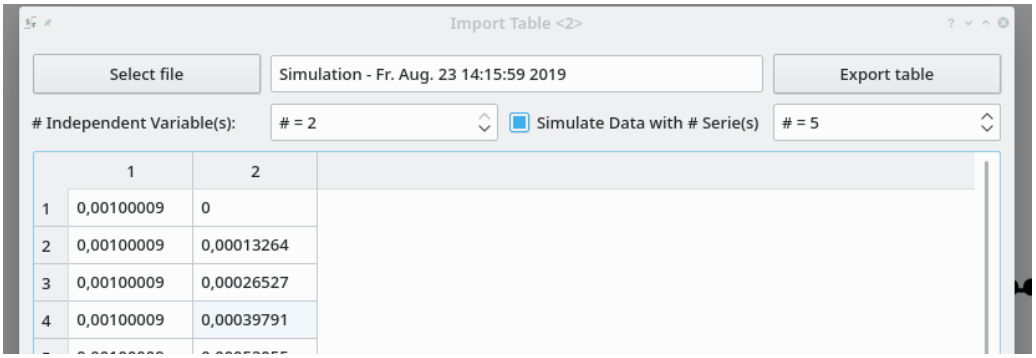
SupraFit files (*.json, *.suprafit) or *.DH files will be loaded automatically without opening the *Import Table* dialog. Importing *.itc file will open the *Thermogram Dialog*. The integration of thermograms is not part of the Quickstart!



(a) Import table with one column of independent variables for ITC or Michaelis-Menten models.



(b) Import table with two columns of independent variables for NMR titrations.



(c) Import table with two columns to simulate NMR titrations with five series.

Figure 2: Import Table dialog in SupraFit. Data can be copied from excel or any other spreadsheet application and pasted by clicking [CTRL-V]. The number of independent variables (columns) have to be specified. Supplement to the analysis of experiments, data can be simulated by checking the **Simulate Data ...** checkbox. The number of series to be simulated is then set to one, but can be increased.

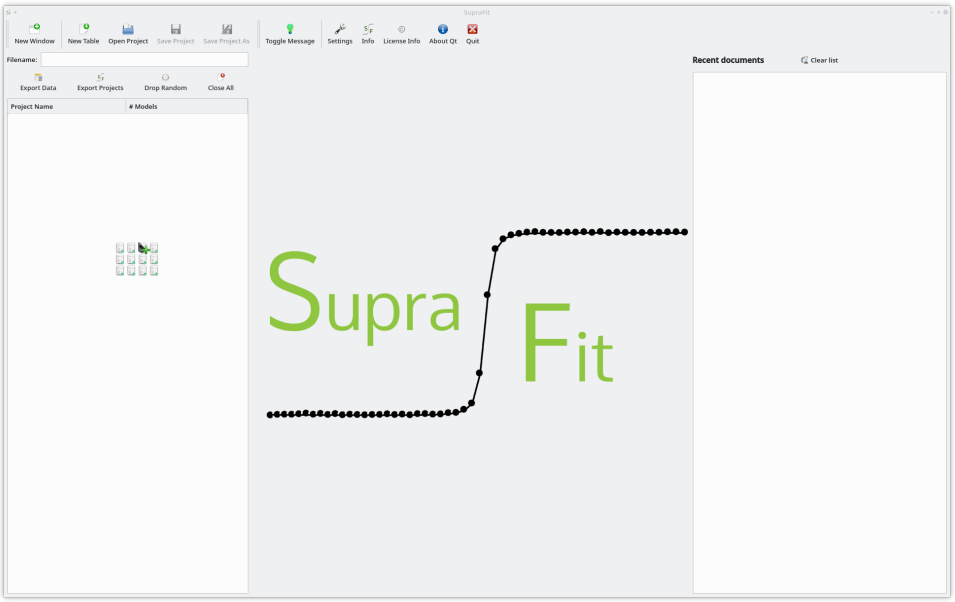


Figure 3: Main window of SupraFit with no open projects. The left white area is the project tree. Drop files into that area to open them.

1.4 Overview Tabulator

A tab is added to the workspace where an overview of the stored data is given, having a bigger text field where any type of comment can be placed - for example the description of experimental conditions. Moreover smaller text field for labels for each data row and scaling factors for the independent data row e.g. concentration (Fig. 4) are provided. More manipulation of the data is currently not supported. Each line in the chart can be hidden - additionally the colour, shape and size of the markers are adjustable. By clicking on the **Single Plot** button any other plot will be hidden and only the chosen one remains in the chart.

Toggling the push button **Hide Datapoints** hides all data points in the charts, without affecting model and error series. Linear regression analysis for the data can be opened by toggling the *Regression* dialog. The assignment of the independent variables can be swapped with the **1<=>2** button, in case their number is two. If there is only one independent variable, the **1<=>2** is hidden and the checkbox **Plot X Values** is shown as long as not model as been added to the workspace: Since SupraFit does not know which values have to be put on the x axis as long as no model has been added, it shows just the number of the experiment (the row number).[6] If the checkbox is checked, the independent variable will be used as values on the x axis.

[6]ITC experiments have the inject volume as independent variable, but instead of the more or less constant inject volume, the ratio of the host and guest concentration is plotted on the x axis. This will not be done before a model has been added, since the same data structure can be used for Michaelis-Menten Kinetics where the independent variable is the x axis. Having NMR titrations, the x axis is typically the ratio of the indepdendent variables instead of the first column.

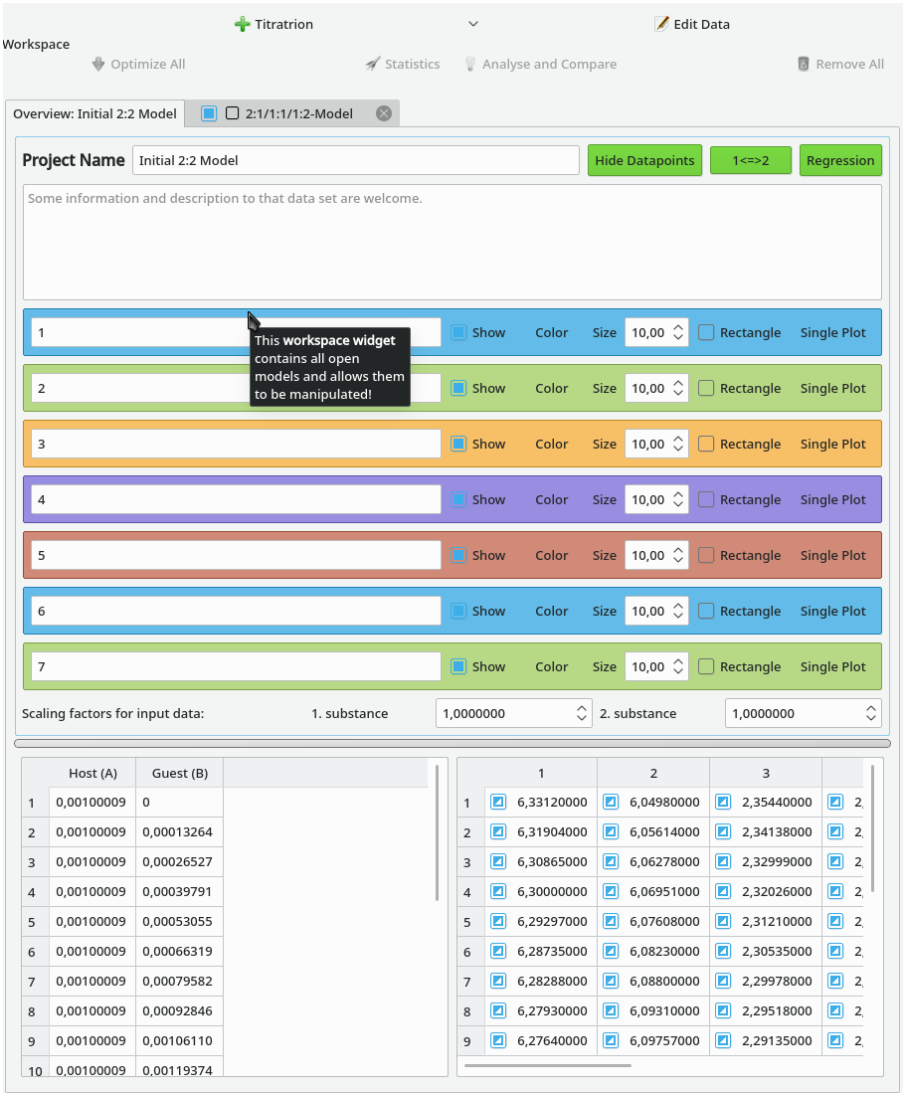


Figure 4: Workspace overview tab with open NMR Titration data.

1.5 Project List and Meta Models

To the left is the project list, where all open projects are stored in a tree view. Double click on any tree item to open the appropriate project and tabulator in the workspace. Right-click *Add Up* or *Substract* to select a file and add up or substract a table: For example after importing a table from a file (table.dat) and then substracting the same file (table.dat) results independent variables filled with zeros. Choosing the same file to be added up results in the original data. Right-click and *Export* saves only the chosen project into a separate file and *Delete* removes the project or model.

Additionally some drag and drop actions are available:

- Drag any model and drop it in the free area below the list to create a new Meta Model (fig. 6a).

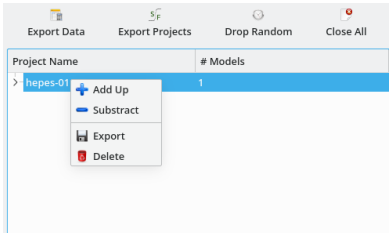
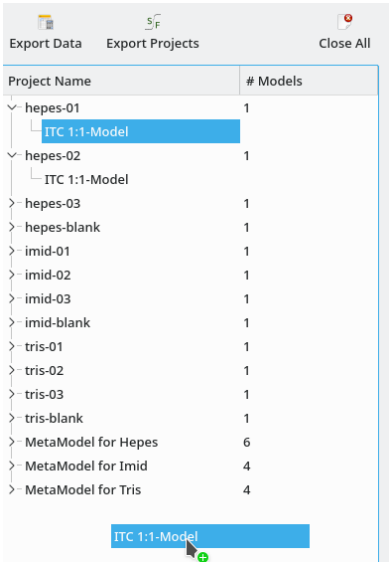
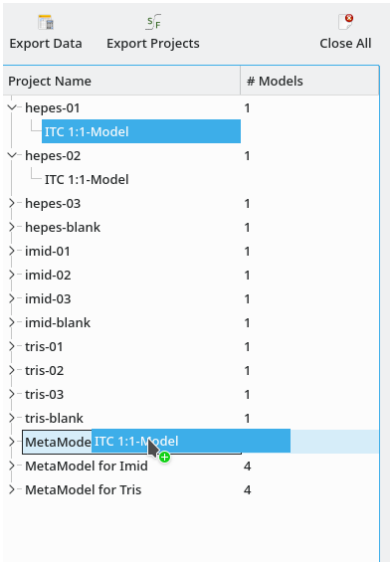


Figure 5: The tree list of open projects and models within the projects.

- Drag any model and drop it into an existing Meta Model to add the selected Model to the Meta Model (fig. 6b).
- Dragging a project into another project (no Meta Models) overrides the system parameter in the destination with those of the source.
- Dragging any model and dropping it into another model of the same type overrides the model's parameters (including models as part of Meta Models).
- Drag data (with or without models) from one open windows of SupraFit and drop it into another window of SupraFit to copy the dragged data (and models)
- Drop a file from the filemanager to open the file



(a) Drop Model on the empty area below the list to create a new Meta Model.



(b) Drop Model on the Meta Model to add it to list of included models.

Figure 6: Drag and Drop actions used to work with Meta Models. The experiments listed were taken from the pytc sample experiments. [7]

1.6 General Charts

On the right side are general charts for models: The upper chart holds the experimental data as well as the recalculated model series, the lower chart displays the absolute errors between the experimental and the calculated data. In all charts in SupraFit the tool button located in the upper right corner. (fig. 7). The **Tools** button allows manipulating the axis' range and labels as well as to export the charts as images in *.png format.

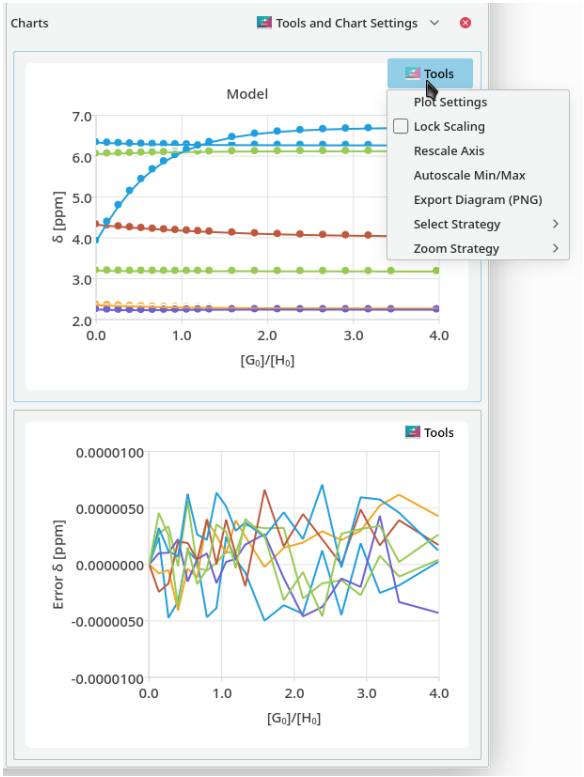


Figure 7: Chart dock with loaded titration curve and the tool button in the upper right corner of every chart.

Generally for every chart in SupraFit: Drawing a rectangle (with the left mouse button) on charts allows zooming into that area, clicking the middle mouse button resets the zoom to the original state.

1.7 Testing Models

1.7.1 NMR and UV/VIS Titrations

To test different models just add a model, e.g. a 1:1 system by choosing it from the **Add Model** menu (Fig 8). NMR and UV/VIS titration can be demanded using first four entries. The subsequent four entries belong to fluorescence titration. These are not fully implemented.

On Windows platform, the separator line with the titles *NMR/UV VIS* and *Fluorescence* is not visible!

This action adds a new tabulator to the workspace where all involved parameters, such as stepwise and cooperative association constants (lgK_{11}, lgK_{21} and lgK_{12}) and shifts, are displayed and can be manipulated easily by changing the value in their appropriate spinboxes. By hovering the mouse on the spinboxes, a tooltip containing the allocation to the model parameters appears. Similar to

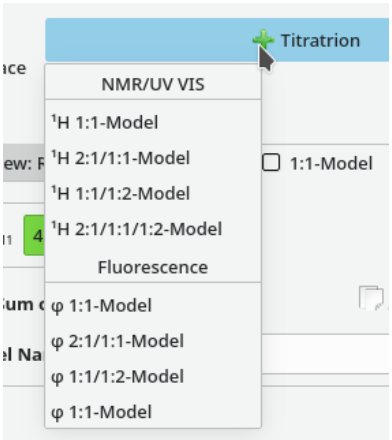


Figure 8: Add different models to the workspace. ^{1}H models can be used for UV/VIS titrations as well.

the titration series, every single recalculated signal series can be selected to be shown in the plot. Pressing **Single Plot** works alike. Optimising the model parameter is being perfomed with the **Fit** button, where all series are used for which the **Include** checkboxes are checked. Unchecking them excludes them from model fitting and statistical analysis.

All open models are being displayed on the charts, but can be deselected by toggling the checkbox (removing the tick) in their tab on the left hand side (fig 10).

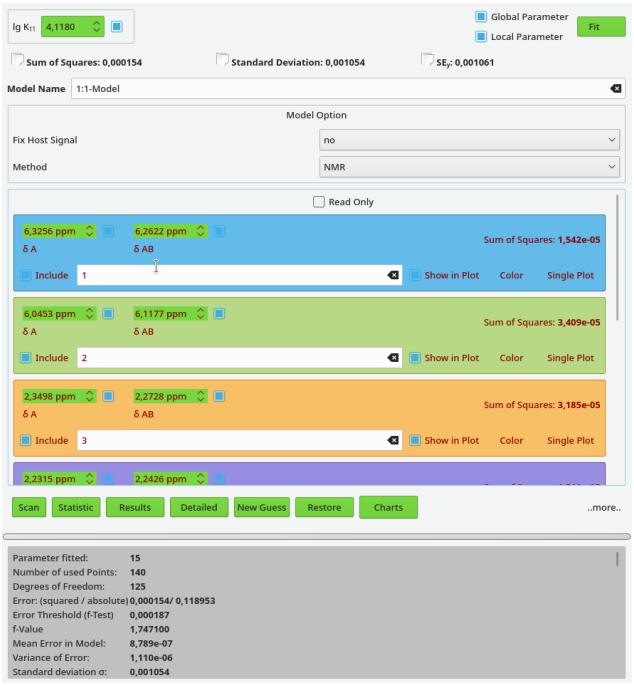


Figure 9: Overview of adjustable model parameters as well as the possible actions in NMR Titration for a 1:1 model.

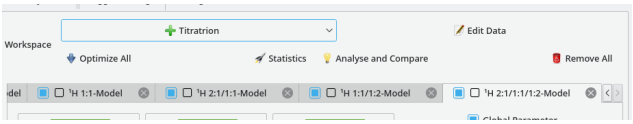


Figure 10: Tabulator widget with models.

1.7.2 ITC

ITC models can be added, if the number of independent variables is set to one. Supplement to the ITC models, one can add Kinetic models (fig. 11).[7]

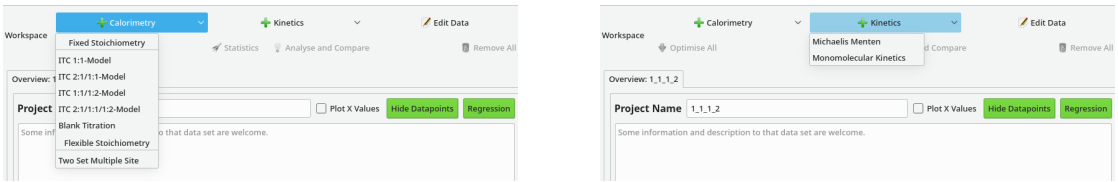


Figure 11: Add different ITC or Kinetic models to the workspace.

ITC models are similar to NMR models, having stability constants and species-specific parameters (the heat of formation) as variables to be fitted. Additionally a linear dilution correction can be

[7] Michaelis-Menten Kinetic is implemented and works as expected, while the Monomolecular Kinetics is just as proof-of-concept implemented. No real experiments have yet been analysed with SupraFit.

included, but not as default value. Another value is *fx*, also known as *n*, the effective concentration of the substance in the cell.

Important parameters that have to be defined for ITC experiments are **system parameters**, like Cell Volume, Cell Concentration and Syringe Concentration. Without these fixed values, no optimisation can be performed.

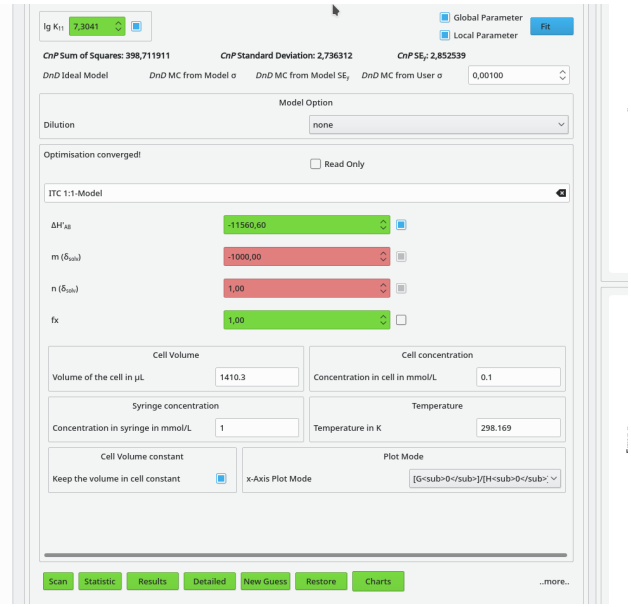


Figure 12: Overview of adjustable model parameters as well as the possible actions in an ITC experiment with a 1:1 model.

1.8 Global Search

Automatic search of various model parameters by performing the optimisation with different initial guesses can be done with **Scan**. A lower and upper bound as well as the increments can be defined for every single parameter. The results will be shown in a table with the resulting *Sum of Squared Errors, (SSE)* as well as the initial and optimised values. Models can be loaded into the current tabulator by double-clicking the left mouse button. Right click opens a menu: Toggling the first action replaces the model parameter with the selected from the table. Choosing the second option adds a new model to the workspace with the selected parameter (fig 13).

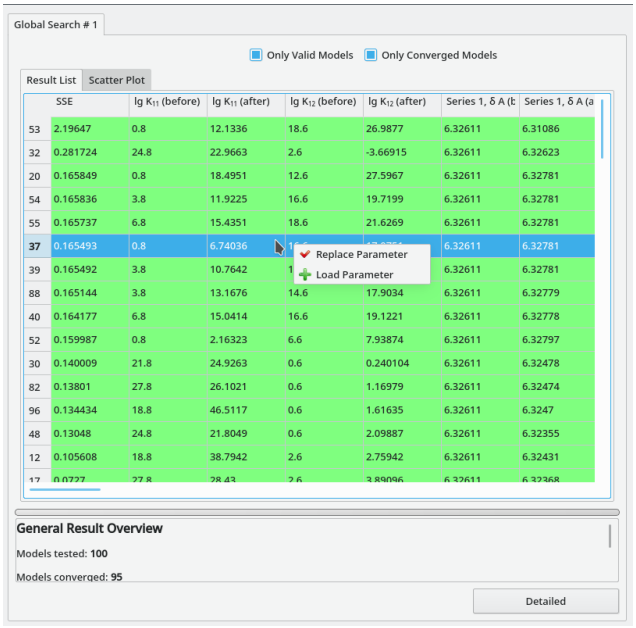


Figure 13: Table of models after global search.

1.9 Model Analysis

After the optimisation has finished (successfully), several post-processing tools are available that help to judge the quality of the obtained models. Some of these approaches analyse the model or model parameters themselves, while other methods need comparison of two or more models. The **Statistics** button in the green toolbox triggers the *Statistics dialog* with three basic approaches, that do not necessarily need comparison of several models. These are:

- Cross Validation and Reduction Analysis
- Monte Carlo Sampling of the (optimised) data set to get confidence intervals for each parameter.
- Error based analysis like Grid Search and Model Comparison.

Some of them have many options to be adjusted, however the default settings are just fine for the start. Weakend Grid Search and Model Comparison are based on the same principle - and being performed completely[8] the confidence intervals obtained are practically identic. Using Monte Carlo Simulation, confidence intervals can be obtained as well - however they do not coincide with the numerical results of Weakend Grid Search.

The **Analyse and Compare** button (fig. 14) holds features to compare several models used in the dataset.

[8]That is not necessarily straightforward.

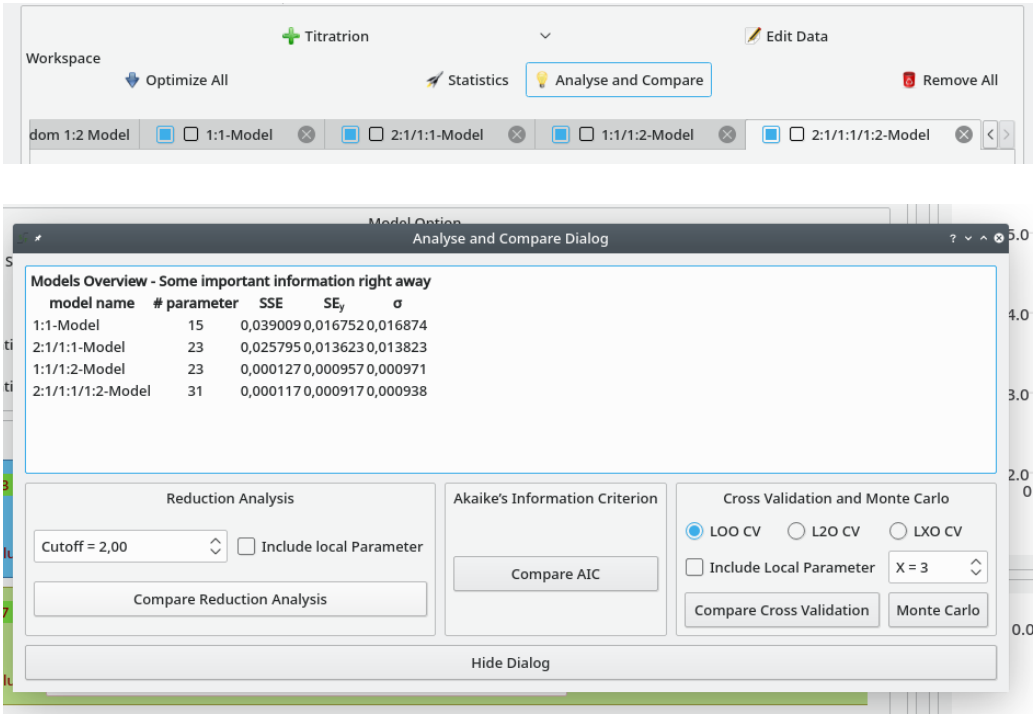


Figure 14: **Analyse and Compare** Button and Dialog along with the **Optimse All** and **Statistic** buttons.

1.10 Known Issues

Apart from regularly appearing and disapperaing bugs and regressions,[9] which should be adressed to Conrad Hübler[10] or at SupraFit@GitHub,[11] the following problems are rooted deeper:

- Statistical results too big can't be stored due to a bug in Qt Json Library: QJsonObject size maximum length 128MB.[12] If that is the case, the least important parts of the data will be dropped. This typically occurs when all intermediate models are to be stored.
- suprafit_cli might crash after long simulations due to a memory leak probably rooted in Qt Json Library as well.

1.11 Citation

If you obtain results with SupraFit, I kindly ask to cite:

[9] This regression is not a mathematical analysis but the (re)-introduction of an error, see wikipedia: https://en.wikipedia.org/wiki/Software_regression.
[10] Conrad.Huebler..at..chemie.tu-freiberg.de
[11] https://github.com/conradhuebler/SupraFit
[12] https://bugreports.qt.io/browse/QTBUG-47629

C. Hübler, conradhuebler/SupraFit: Version 1.7.0 **2019**, Zenodo.
http://doi.org/10.5281/zenodo.3364570

A detailed paper for SupraFit and some adopted as well as new introduced methods will follow.

References

[1] M. J. Hynes, *J. Chem. Soc. Dalton Trans.* **1993**, 311–312.

[2] C. Frassineti, S. Ghelli, P. Gans, A. Sabatini, M. S. Moruzzi, A. Vacca, *Anal. Biochem.* **1995**, *231*, 374–382.

[3] P. Thordarson, *Techniques in Supramolecular Chemistry : From Molecules to Nanomaterials*, (Eds.: P. A. Gale, J. W. Steed), WILEY-VCH Verlag GmbH & Co. KGaA, **2012**.

[4] P. Thordarson, supramolecular.org, http://supramolecular.org, **2019**.

[5] P. Thordarson, Open Data Fit, http://opendatafit.org/, **2019**.

[6] T. Instruments, NanoAnalyze: Texas Instruments, 10.07.2018, http://www.tainstruments.com.

[7] H. Duvvuri, L. C. Wheeler, M. J. Harms, *Biochemistry* **2018**, *57*, PMID: 29620867, 2578–2583.

[8] P. Thordarson, *Chem. Soc. Rev.* **2011**, *40*, 1305–1323.

[9] Qt- Toolkit 5.13: https://www.qt.io/ 7.08.2019.

[10] G. Guennebaud, B. Jacob, et al., Eigen v3, http://eigen.tuxfamily.org, **2010**.

[11] A. Vacca, O. Francesconi, S. Roelens, *Chem. Rec.* **2012**, *12*, 544–566.

[12] L. K. von Krbek, C. A. Schalley, P. Thordarson, *Chem. Soc. Rev.* **2017**, *46*, 2622–2637.

[13] H. Motulsky, A. Christopoulos, *Fitting Models to Biological Data using Linear and Nonlinear Regression. A practical guide to curve fitting*. GraphPad Software Inc., San Diego CA, **2003**.

[14] J. M. Beechem in *Numerical Computer Methods*, Methods in Enzymology, Academic Press, **1992**, pp. 37–54.

[15] H. Akaike, *Trans. Autom. Control* **1974**, *19*, 716–723.

[16] N. Sugiura, *Commun. Stat. - Theory Methods* **1978**, *7*, 13–26.

[17] GNU General Public License: http://www.gnu.org/licenses/gpl.html 11.10.2018.

[18] K. Hirose in *Analytical Methods in Supramolecular Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, **2012**, Chapter 2, pp. 27–66.

[19] E. Freire, A. Schön, A. Velazquez-Campoy, *Methods in enzymology* **2009**, *455*, 127–155.

[20] A. Velazquez-Campoy, *J Therm. Anal. Calorim.* **2015**, *122*, 1477–1483.

[21] M. W. Freyer, E. A. Lewis, *Methods in cell biology* **2008**, *84*, 79–113.

[22] O. Francesconi, C. Nativi, G. Gabrielli, M. Gentili, M. Palchetti, B. Bonora, S. Roelens, *Chem. Eur. J.* **2013**, *19*, 11742–11752.

[23] K. Levenberg, *Quarterly of applied mathematics* **1944**, *2*, 164–168.

[24] D. W. Marquardt, *SIAM Journal on Applied Mathematics* **1963**, *11*, 431–441.

[25] C. P. Robert, G. Casella, *Monte Carlo Statistical Methods*, 2nd ed., Springer Verlag, **2004**.

[26] M. Matsumoto, T. Nishimura, *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **1998**, *8*, 3–30.

[27] G. Kemmer, S. Keller, *Nature protocols* **2010**, *5*, 267–281.