SupraFit - An Open source Qt based fitting application to determine stability constants from titration experiments

Conrad Hübler*[a]

A novel application to determine stability constants from supramolecular titration experiments is presented. The focus lies on NMR titration and ITC experiments for pure 1:1 systems, as well as mixed 2:1/1:1, 1:1/1:2 and 2:1/1:1/1:2 systems. *SupraFit* provides global and local fitting and a global search tool. Statistical methods are implemented and can be applied to analyse the results of nonlinear regression. Monte Carlo simulations, combined with the percentile methods and F-Test approaches to calculate confidence intervals are supported. The implemented statistical approaches are illustrated and discussed on model functions. All methods are accessible through an intuitive user interface, providing charts for all (kind of) data produced. *SupraFit* is written in C++, using the Qt Toolkit for the Graphical User Interface (GUI) and the Eigen library for nonlinear regression and is released under the GNU Public License (GPL).

1 Introduction

After the work of Pederson, Lehn and Cram in the second half of the 20th century (nobel prize in 1987 “for their development and use of molecules with structure-specific interactions of high selectivity”), supramolecular chemistry has become a popular field of research. The experimental determination of association constants utilising supramolecular titration experiments plays a big role in the analytical zoo of this research area. Several software packages have been written in the last three decades, each having its own strength and weaknesses. In times of open science, open data and open source software, some of these older software solutions might be considered as not state-of-the-art. The most recent tool for supramolecular titration experiments has been developed by the group of Thordarson and is available via www.supramolecular.org (last checked 17.01.2022) as online service. As a matter of taste, someone could prefer online tools to offline applications and vice verse. As an alternative to the older offline applications and as well as to the online tools, a newly written software package for supramolecular titration experiments called *SupraFit* is reported.

SupraFit is written in C++ utilising the Qt Software Development Toolkit[1] and the Eigen Library.[2] *SupraFit* is mainly developed for NMR titration and ITC experiments, providing methods to globally and locally analyse 1:1, 2:1/1:1 and 1:1/1:2 complexes out of the box. Fully statistical analysis based on Monte Carlo simulation and F-Test approaches with a good scaling on multicore systems are implemented as well as an intuitive user interface to deal with several models on single data sets. Due to being open source, own models can be implemented in the source code, with all functionality eg. statistical analysis being provided for the new models.

2 Software

Several packages already exist for the analysis of NMR titrations or ITC data, some of them did not receive updates or improvements recently. Additionally, these programs may provide statistical analysis, which are not always comparable to each other as they are based on different theories. A third point is the advantage of software to run on different operating systems (OS) or even being independent of an OS, although Windows systems dominate the PC market.

In the last decade, the idea of open source software, as well as open data has evolved, and more scientific software is not only freely usable but the source code is published under the terms of an open source licences, such as *GPL* or[3] *MIT*.[4] In contrast to SupraFit, the available open source programs are mainly focused on computational chemistry and chemoinformatics.[5–8]

Some common tools used to analyse supramolecular titration experiments will be listed in the next section, however without any claim to completeness.

2.1 NMR Titration

WinEQNMR, initially a DOS program called *EQNMR*, has been written by M. J. Hynes[9] and is available for Windows systems. *WinEQNMR* provides methods for protonation equilibria, hydrolysis of metal ions or stability of metal complexes. An archive containing the binaries was freely downloadable at http://www.nuigalway.ie/chem/Mike/wineqnmr.htm. However the website is not available any more, but can be accessed via the wayback machine (https://web.archive.org/web/20210518005317-/http://www.nuigalway.ie/chem/Mike/wineqnmr.htm). The password protected archive containing the program is not available via the wayback machine service.

HypNMR[10] is part of the Hyperquad software package devel-

[a] *Institut für Organische Chemie, Technische Universität Bergakademie Freiberg, Leipziger Strasse 29, Freiberg, Germany. E-mail: Conrad.Huebler@gmx.net*

† Electronic Supplementary Information (ESI) available: Selected images with higher resolution (Fig. S1 - S4). Input data for simulated 1:1/1:2 model. Selected images with higher resolution (Fig. S5). Calculated intersection in Mole Ratio plot (Table S1). Selected images with higher resolution (Fig. S6). Representation of boxplots from Monte Carlo simulation for NMR titration (Fig. S7 - S14). Comparison of initial guessed and least-squared estimated parameters in ITC experiments (Table S2). Isotherms and dilution experiments (Fig. S15). Estimated parameters and confidence interval with different dilution strategies (Table S3 - S5, Fig. S16 - S20).

oped Sabatini, Vacca and coworkers, providing tools for different methods such as NMR titration, ITC and spectrophotometry. *HypNMR* runs on Windows system and information on how to obtain the software are available upon request.[‡] The most recent version according to their website (http://www.hyperquad.co.uk/hypnmr.htm, last checked 17.01.2022) is HypNMR2018, without pointing out the differences to the older versions.

M. Maeder and P. King founded *Jplus Consulting* in 2009 and provide a software packages called ***ReactLab*** to analyse and simulate for example equilibrium titrations and kinetics. The software is based on a combination of *MatLab* and *Excel* and is available for purchase. More information can be found on their official web site: (https://jplusconsulting.com (last checked 17.01.2022).

Open Data Fit[11] is a collection of online services provided by P. Thordarson, where titration data can be analysed. The service can be accessed at http://opendatafit.org (last checked 17.01.2022). For now, supramolecular experiments[12] and a demo version for cell viability[13] are available. The kinetics service is under construction.[14] BindFit, the part focusing on supramolecular titration, was initially provided as free *MatLab* scripts included in the tutorial review by Thordarson 2011.[15] The latest version supports analysis of NMR and UV/VIS titration of typical 1:1, 2:1 and 1:2 systems, with the python source code being available at https://github.com/echus/supramolecular-apps (last checked 17.01.2022). New features such as Monte Carlo simulation based statistics are announced for future versions.

2.2 ITC

NanoAnalyze is available from *TA Instruments*, that assemble and sell instruments for several analysis (thermal, microcalorimetric and rheologic analysis). *NanoAnalyze* is freely available for Windows systems, provides several binding models, analysis of thermograms and statistics based on Monte Carlo simulations. It can be obtained from their website https://www.tainstruments.com/itcrun-dscrun-nanoanalyze-software (last checked 17.01.2022).

Harms et al.[16] released ***pytc*** (**py**thon **itc**) as open source software, built on top of python3 to analyse ITC data, having the most important binding models already implemented. The project is hosted on GitHub: https://github.com/harmslab/pytc (last checked 17.01.2022). Since it is written in *python3*, other models can easily be added. Statistical methods like F-Test or Information Criterion[17–20] methods are implemented and can be used to determine the performance of the models. An graphical user interface using *PyQt5* can be downloaded separately at https://github.com/harmslab/pytc-gui (last checked 17.01.2022).

SEDFIT and ***SEDPHAT*** form a program package to globally analyse ITC data (gITC), with powerful statistical analysis based on Monte Carlo simulations or the F-Test approach.[21] It is freely available at https://sedfitsedphat.nibib.nih.gov/software/default.aspx (last checked 17.01.2022), however other systems apart from windows are not supported. Thermogram analysis can be performed with *NITPIC* (http://biophysics.swmed.edu/MBR/software.html last checked 17.01.2022) from Keller et al.[22] and then imported into *SEDFIT*.

3 Supramolecular Titrations

The theory of complexation and supramolecular titration is already reviewed in articles by Thordarson,[15,23] as well as in text books like Analytical Methods in Supramolecular Chemistry[24] but the main aspects will be summarised here:

3.1 General Approach

Starting from the general mass balance equations (eq. 1 and 2) for a two-component system, the relationship between the concentration of two components [A] and [B] can be described through the cumulative stability constants (eq. 3). For example individual stability constants for a system with two complex species A_aB_b defined with $a = b = 1$ and $a = 2, b = 1$ read as in equation 4.

$$[A]_0 = \sum_{\substack{a=0\\b=0}}^{l,m} a\beta_{ab}[A]^a[B]^b \tag{1}$$

$$[B]_0 = \sum_{\substack{a=0\\b=0}}^{l,m} b\beta_{ab}[A]^a[B]^b \tag{2}$$

$$\beta_{ab} = \prod_{\substack{a=0\\b=0}}^{l,m} K_{ab} \tag{3}$$

$$K_{11} = \frac{[AB]}{[A][B]} \qquad K_{21} = \frac{[A_2B]}{[A][AB]} \tag{4}$$

Depending on the values for l and m, e.g. the stoichiometry of molecules of A and B that are involved in forming the complex, different systems can be described. *SupraFit* reports all stability constants as individual logarithmic constants lgK, in contrast to other software that may report them as plain stability constants K in M^{-1} or as cumulative constants β.

3.2 Determining stability constants

The determination of association constants with titration experiments is based on the idea, that each component influences the response signal: Assuming a linear relationship between the amount of species and the response signal, equation 5 can be formed, where each component X_i contributes to the overall signal y by a factor Y_i.

$$y = \sum_i Y_i[X]_i \tag{5}$$

3.2.1 NMR Titration

Upon performing ^{1}H-NMR titration, the chemical shift of specific protons bound to X (eg. receptor) changes during complexation due to non-covalent interactions with another component. Depending on the kinetics of the complex formation, fast and slow

[‡] See http://www.hyperquad.co.uk/index.htm for more information on how to obtain parts of the *Hyperquad* software products.

exchange can be observed. *SupraFit*, as most of the other applications, can only handle fast exchange, where the observed signal is the weighted average of all signals of the specific proton in the components, e.g. the shift of a proton assigned to the isolated receptor and one to the complex.

Since the relative change of the chemical shifts is of interest, it is defined as the ratio of each component to the reference component: in following case using the first component. Equation 5 reads for NMR titration as follows:

$$\Delta\delta = \sum_i \delta_i \frac{[X]_i}{[X]_0} \tag{6}$$

On the other hand, for the slow exchange, for each component a signal for the specific proton can be observed, where the intensity is related to the amount of the species.[25]

3.2.2 UV/VIS Titration

In the UV/VIS titration, the overall absorbance is the sum of the individual extinction coefficient ε_i multiplied by concentration of each component. The equation holds true for low concentrations that fulfill Lambert-Beers Law.

$$A_{abs} = \sum_i A_{abs,i} = \sum_i \varepsilon_i [X]_i \tag{7}$$

3.2.3 ITC

General aspects

The basic part of isothermal titration calorimetry is the observation of the change of heat due to a complex formation in a reaction cell while keeping the temperature constant. The guest component B is sequentially added to a solution of the host component A. Details on that method can be found in literature of Freire,[26,27] in Analytical Methods of Supramolecular Chemistry[28] by Schmidtchen, as well as in reviews by Thordarson.[15,29]

The basic ITC equation 8 describes a sum over all formed complex species multiplied with corresponding heat of formation. In contrast to NMR and UV/VIS titrations, the pure host signal does not contribute to the observed heat. At the current state, *SupraFit* only makes use of models, that are of fixed stoichiometry and equal to the well known NMR titration models, that are summarised in section 3.3. Furthermore, *SupraFit* handles titration experiments with both, a fixed-volume set up as well as a set up with variable volume.

$$Q = V \sum_i \Delta H_i [X]_i \tag{8}$$

Handling dilution effects

Since upon each injection of B the concentration of B itself changes, an amount of signal can be lead back to a heat of dilution (Q_d), that cannot be neglected. Assuming a linear relationship between the concentration of B and the response heat signal, one can use equation 9 to add blank effects to the experiment (eq. 10), as done for example in pytc.[16]

$$Q_{d,i} = m_\delta [B_i] + n_\delta \tag{9}$$

$$Q = V \sum_i \Delta H_i [X]_i + m_\delta [B_i] + n_\delta \tag{10}$$

As a consequence, different approaches to deal with the dilution can be realised using SupraFit:

1. Using equation 10, two parameters are introduced (m_δ and n_δ) and fitted alongside with the stability constants and the heat of formation to the experimental titration curve. An additional blank experiment does not have to be performed.

2. The two blank parameters (m_δ and n_δ) are obtained from an independent dilution experiment and are added as fixed terms to equation 10.

3. The result of the independent dilution experiment is subtracted from the titration experiment and which is used to fit the parameters in equation 8 afterwards.

4. The blank parameters are fitted to a blank experiment and the titration simulaneously, while the stability constants and the heat of formation are fitted to the titration experiment only (eq. 10).

Thermogram handling

SupraFit provides ready-to-use thermogram integration functions with elementary baseline corrections for *.itc and plain thermogram files consisting of columns with time and heat per time, respectively. The baseline is separately calculated for each peak as a linear function, where the integration range can be adjusted manually. In case of very unregular baselines, different software packages may be more sufficient, such as NITPIC or software provided by the hardware supplier. After integration using third party software, the plain data can be processed with *SupraFit*.

3.3 1:1 Model

The simplest form of complexes with two components are the 1:1 complexes ($a = 1$, $b = 1$), which are formed according to equation 11. K_{11} denotes the step-wise complex formation constant. The approach is sketched in Appendix B, resulting in equation 12.

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB} \qquad K_{11} = \frac{[AB]}{[A][B]} \tag{11}$$

$$\begin{aligned} 0 = K_{11}[AB]^2 - [AB](K_{11}[A]_0 + \\ + K_{11}[B]_0 + 1) + K_{11}[A]_0[B]_0 \end{aligned} \tag{12}$$

Using the solution of $[AB]$ from the quadratic equation 12 all remaining concentrations can be calculated according to the mass-balance equation. The resulting equations for 1:1 models used in *SupraFit* are summarised in Table 1 with only the shifts of the host and the complex are taken into account. Signals of component B are ignored. For UV/VIS this holds true if the component is not UV/VIS active at the selected wave length.

Table 1 Equations used in 1:1 models

Method	Equation
NMR	$\delta_{calc} = \delta_A \frac{[A]}{[A]_0} + \delta_{AB} \frac{[AB]}{[A]_0}$
UV/VIS	$A_{Abs,calc} = \varepsilon_A [A] + \varepsilon_{AB} [AB]$
ITC	$Q_i = V \left([AB]_i - [AB]_{i-1} \cdot (1 - \frac{v}{V}) \right) \cdot \Delta H_{AB}$

Table 2 Equations used in 2:1/1:1 models

Method	Equation
NMR	$\delta_{calc} = \delta_A \frac{[A]}{[A]_0} + \delta_{AB} \frac{[AB]}{[A]_0} + 2\delta_{A_2B} \frac{[A_2B]}{[A]_0}$
UV/VIS	$A_{abs,calc} = \varepsilon_A [A] + \varepsilon_{AB} [AB] + 2\varepsilon_{A_2B} [A_2B]$
ITC	$Q_i = V(\left([AB]_i - [AB]_{i-1} \cdot (1 - \frac{v}{V}) \right) \cdot \Delta H_{AB} +$ $+ \left([A_2B]_i - [A_2B]_{i-1} \cdot (1 - \frac{v}{V}) \right) \cdot \Delta H_{A_2B})$

3.4 2:1/1:1 Model

A model of 2:1/1:1 stoichiometry is defined through the following relationship:

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB} \qquad K_{11} = \frac{[AB]}{[A][B]} \tag{13}$$

$$\mathrm{A} + \mathrm{AB} \rightleftharpoons \mathrm{A_2B} \qquad K_{21} = \frac{[A_2B]}{[A][AB]} \tag{14}$$

The stepwise stability constants K_{11} and K_{21} combine to the cumulative association constants as follows:

$$K_{11}K_{21} = \frac{[AB]}{[A][B]} \frac{[A_2B]}{[A][AB]} = \frac{[A_2B]}{[A]^2[A]} = \beta_{21} \tag{15}$$

The solution for the concentration of A is given in equation 47 in the appendix.[15] The corresponding equations to describe a 2:1/1:1 model used within *SupraFit* are summarised in Table 2, with the guest molecule being silent. In case of ITC experiments, 2:1/1:1 are not used regularly, but have already been reported.[30]

3.5 1:1/1:2 Model

The 1:1/1:2 system is defined through following law of mass action:

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB} \qquad K_{11} = \frac{[AB]}{[A][B]} \tag{16}$$

$$\mathrm{AB} + \mathrm{B} \rightleftharpoons \mathrm{AB_2} \qquad K_{12} = \frac{[AB_2]}{[AB][B]} \tag{17}$$

The concentration of unbound guest can be calculated analogously to the 2:1/1:1 systems using equation 50,[15] where the free host concentration can be determined using the mass-balance equations for 1:1/1:2 system.

Having the free and complexed host concentrations, the signals are calculated in *SupraFit* using the equations in Table 3, with the guest molecule being silent.

Table 3 Equations used in 1:1/1:2 models

Method	Equation
NMR	$\delta_{calc} = \delta_A \frac{[A]}{[A]_0} + \delta_{AB} \frac{[AB]}{[A]_0} + \delta_{AB_2} \frac{[AB_2]}{[A]_0}$
UV/VIS	$A_{abs,calc} = \varepsilon_A [A] + \varepsilon_{AB} [AB] + \varepsilon_{AB_2} [AB_2]$
ITC	$Q_i = V(\left([AB]_i - [AB]_{i-1} \cdot (1 - \frac{v}{V}) \right) \cdot \Delta H_{AB} +$ $+ \left([AB_2]_i - [AB_2]_{i-1} \cdot (1 - \frac{v}{V}) \right) \cdot \Delta H_{AB_2})$

3.6 2:1/1:1/1:2 Model

The last titration model implemented in *SupraFit* is the mixed model with 2:1, 1:1 and 1:2 species.

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB} \qquad K_{11} = \frac{[AB]}{[A][B]} \tag{18}$$

$$\mathrm{AB} + \mathrm{B} \rightleftharpoons \mathrm{AB_2} \qquad K_{12} = \frac{[AB_2]}{[AB][B]} \tag{19}$$

$$\mathrm{AB} + \mathrm{A} \rightleftharpoons \mathrm{A_2B} \qquad K_{21} = \frac{[A_2B]}{[AB][A]} \tag{20}$$

The solution of this system is defined by the mass-balance equation

$$[A]_0 = [A] + \beta_{11}[A][B] + \beta_{12}[A][B]^2 + 2\beta_{21}[A]^2[B] \tag{21}$$

$$[B]_0 = [B] + \beta_{11}[A][B] + 2\beta_{12}[A][B]^2 + \beta_{21}[A]^2[B] \tag{22}$$

The mass balance equation can be simplified and reads as:

$$[A]([B]) = (2\beta_{21}[B]) \cdot [A]^2 + (\beta_{12}[B]^2 + K_{11}[B] + 1) \cdot [A] - [A]_0 \tag{23}$$

$$[B]([A]) = (2\beta_{12}[A]) \cdot [B]^2 + (\beta_{21}[A]^2 + K_{11}[A] + 1) \cdot [B] - [B]_0 \tag{24}$$

The solution to this equilibrium system is obtained using an iterative procedure: The initial concentrations are guessed as

$$[A] = min([A]_0, [B]_0)/10$$

$$[B] = B([A]) \text{ (according to eq. 24)}$$

followed by the calculation of $[A]$ and $[B]$ with then equation 23 and 24. The calculations are repeated until the change in the equilibrium concentrations reaches a threshold. Alternatively to this algorithm, methods to solve any equilibria system based on a Gauss-Newton optimisation have been published.[31] A Levenberg-Marquardt optimisation has been tested in *SupraFit*, but was disabled.[§]

[§] During Monte Carlo simulations the Levenberg-Marquardt optimisation was not as efficient as the approach described above. However, a detailed benchmark was not prepared.

Table 4 Equations used in 2:1/1:1/1:2 models

Method	Equation
NMR	$\delta_{calc} = \delta_A \frac{[A]}{[A]_0} + \delta_{AB}\frac{[AB]}{[A]_0} + 2\delta_{A_2B}\frac{[A_2B]}{[A]_0} + \delta_{AB_2}\frac{[AB_2]}{[A]_0}$
UV/VIS	$A_{abs,calc} = \varepsilon_A[A] + \varepsilon_{AB}[AB] + 2\varepsilon_{A_2B}[A_2B] + \varepsilon_{AB_2}[AB_2]$
ITC	$\begin{aligned} Q_i = V(& \left([AB]_i - [AB]_{i-1}\cdot(1-\frac{v}{V})\right)\cdot\Delta H_{AB} + \\ & +\left([A_2B]_i - [A_2B]_{i-1}\cdot(1-\frac{v}{V})\right)\cdot\Delta H_{A_2B}) \\ & +\left([AB_2]_i - [AB_2]_{i-1}\cdot(1-\frac{v}{V})\right)\cdot\Delta H_{AB_2}) \end{aligned}$

Cooperativity

Cooperative effects describe increasing or decreasing step-wise bindings constants in multi-step systems and have been discussed in the literature.[29,32,33] Following the notation of Thordarson,[11,15,29] four different types can be distinguished: full, noncooperative, additive and statistical. These models can be applied to 2:1 and 1:2 complex species in the mixed models in *SupraFit*. The different kinds of relationship that can be set up in the model options are summarised in Table 5.

Table 5 Different cooperative binding models define the relationship of the estimated model parameters. The relationships are taken from Hibbert and Thordarson, 2016.[11] K_2 refers to either K_{12} or K_{21}, depending on the stoichiometry of the complex. Similar, $\delta_{\Delta 2}$ refers to the signal of either the 2:1 or 1:2 species, whereas $\delta_{\Delta 1}$ denotes the 1:1 species

model	K	δ
full	$K_1 \neq 4K_2$	$\delta_{\Delta 2} \neq \delta_{\Delta 1}$
noncooperative	$K_1 = 4K_2$	$\delta_{\Delta 2} \neq \delta_{\Delta 1}$
additive	$K_1 \neq 4K_2$	$\delta_{\Delta 2} = \delta_{\Delta 1}$
statistical	$K_1 = 4K_2$	$\delta_{\Delta 2} = \delta_{\Delta 1}$

3.7 Michaelis-Menten Theory

Michaelis-Menten theory is usually used to describe how the rate r of an enzymatic reaction, that transforms a substrate S to a product P (eq. 25), depends on the amount of substrate S_0.[34]

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES}$$

$$\mathrm{ES} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{E} + \mathrm{P} \tag{25}$$

The rate is defined as

$$r = \frac{v_{max}\cdot S}{K_M + S} \tag{26}$$

At high concentrations of S, the rate r tends towards v_{max}. A linearised form of the Michaelis-Menten equations, the Lineweaver-Burke form (eq. 27), is usually used to determine K_M and v_{max}.

$$\frac{1}{r} = \frac{K_M}{v_{max}}\frac{1}{S} + \frac{1}{v_{max}} \tag{27}$$

SupraFit provides a model to determine K_M and v_{max} using nonlinear regression. The starting guess is calculated using eq. 27.

3.8 Nonlinear least-squares regression

The set of unknown parameters $\underline{\theta}$, that are used to describe the relation of the independent data x and the experimental data y_{exp} (eq. 28), have to be adjusted to minimise the sum of squared errors (SSE, eq. 29). In case of NMR titrations $\underline{\theta}$ corresponds to the stability constants and chemical shifts each component, x to the concentrations and y to the observed chemical shifts. In connection with ITC experiments $\underline{\theta}$ refers again to the the stability constants as well as the heat of formation and optional to the dilution parameters. The integrated peaks of the a thermogram form y and the concentrations remain to be the independent parameters x.

For the nonlinear problem, the Levenberg-Marquardt Algorithm[35,36] as implemented in Eigen, is used.

$$y_{calc,i} = f(\theta, x_i) + e_i \tag{28}$$

$$SSE = \sum_i (y_{exp,i} - y_{calc,i})^2 = \sum_i e_i^2 \rightarrow 0 \tag{29}$$

$y_{exp,i}$ denotes the experimentally observed value at i, $y_{calc,i}$ the estimation of the observed value according to the model parameter and e_i the residual at each data point. The parameters θ are henceforth referred as to $\hat{\theta}$ in case they are the best-fit parameters after least-squares optimisation. Characterisation of the fit can be realised using the standard deviation of the residuals σ_{fit} (eq. 30), SE_y (eq. 31) and χ^2 (eq. 32):[15]

$$\sigma_{fit} = \frac{\sum_i e_i^2}{N-1} \tag{30}$$

$$SE_y = \frac{\sum_i e_i^2}{N-k} \tag{31}$$

$$\chi^2 = \frac{\sum_i e_i^2}{N-k-1} \tag{32}$$

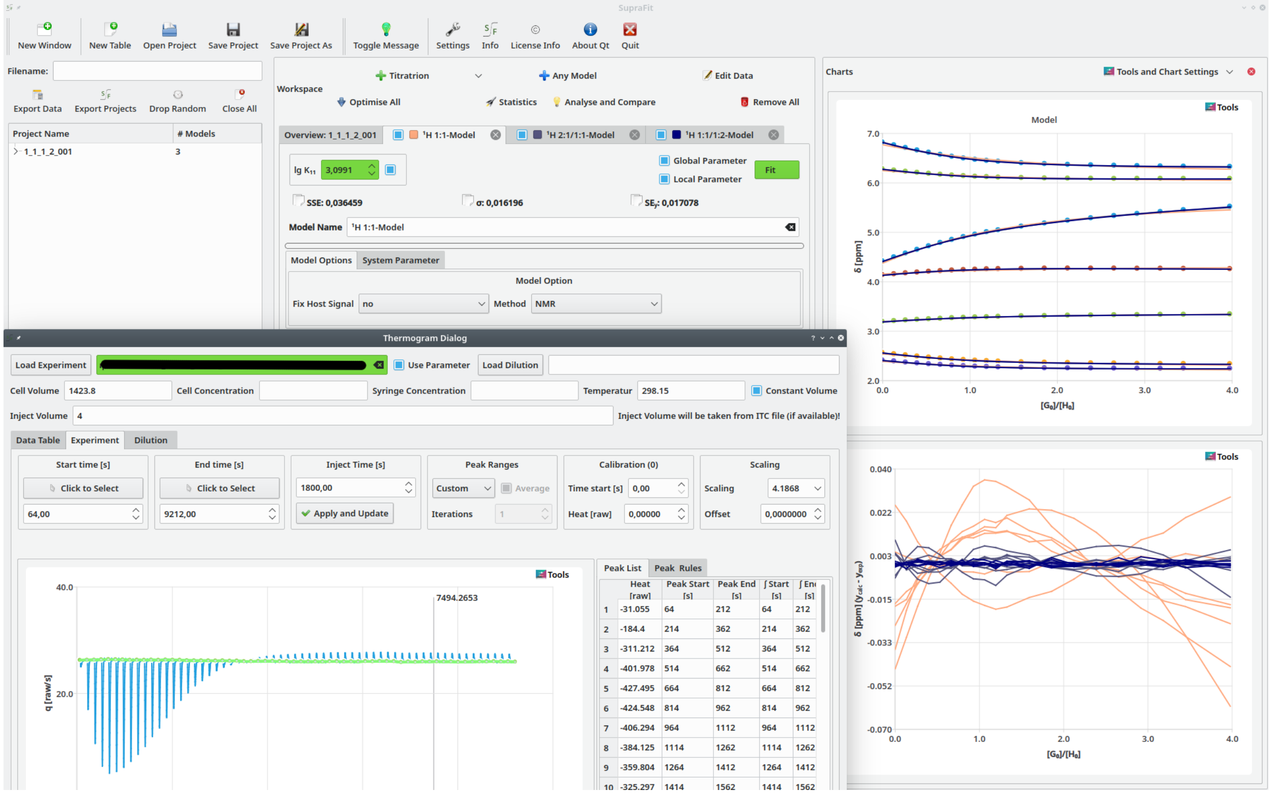
SE_y is the corrected standard deviation with respect to the number of parameters (k) in the applied model.

4 Features

4.1 General

An introduction to *SupraFit* is not reported in that article, it can be found in the *SupraFit Quickstart*,[37] however the main aspects will be summarised: The *SupraFit* package contains two binaries, the *suprafit.exe* binary providing the graphical user interface (GUI) and *suprafit_cli.exe* providing command line interface. The GUI comes with all basic functionalities for loading and saving data sets as well as thermogram integration in case of ITC experiments. Most of the results obtained with *SupraFit* are provided as adjustable charts and text information, where the diagrams can be exported to *.png files. Many charts presented in this article were exported directly from *SupraFit*, the remaining charts, mainly the boxplots, are LaTeX and Ti*k*Z based. A screenshot of the main window can be found in Fig. 1.

SupraFit reads simple Table files as well as *.itc files. For the later, the thermogram import is straight forward. Additionally, data simulation and basic experimental planning are available



Filename:

Project Name	# Models
1_1_1_2_001	3

Overview: 1_1_1_2_001 | ¹H 1:1-Model | ¹H 2:1/1:1-Model | ¹H 1:1/1:2-Model

lg K_{11} 3,0991

☐ Global Parameter

☐ Local Parameter

Fit

SSE: 0,036459

σ: 0,016196

SE_y: 0,017078

Model Name ¹H 1:1-Model

Model Options | System Parameter

Model Option

Fix Host Signal: no

Method: NMR

[redacted]

☐ Use Parameter

Cell Volume 1423.8

Cell Concentration

Syringe Concentration

Temperatur 298.15

☐ Constant Volume

Inject Volume 4

Inject Volume will be taken from ITC file (if available)!

Start time [s]: 64,00

End time [s]: 9212,00

Inject Time [s]: 1800,00

Peak Ranges: Custom ☐ Average

Iterations 1

Calibration (0)

Time start [s] 0,00

Heat [raw] 0,00000

Scaling

Scaling 4.1868

Offset 0,0000000

	Heat [raw]	Peak Start [s]	Peak End [s]	∫ Start [s]	∫ En [s]
1	-31.055	64	212	64	212
2	-184.4	214	362	214	362
3	-311.212	364	512	364	512
4	-401.978	514	662	514	662
5	-427.495	664	812	664	812
6	-424.548	814	962	814	962
7	-406.294	964	1112	964	1112
8	-384.125	1114	1262	1114	1262
9	-359.804	1264	1412	1264	1412
10	-325.297	1414	1562	1414	1562

Fig. 1 Screenshot of the main window with the dialog box to import thermograms open.

with the current functions. More details on the usage of *SupraFit* are available in the quickstart, that can be downloaded on the GitHub webpage at https://github.com/conradhuebler/SupraFit.

4.2 Technical aspects and implementation

SupraFit is written in C++ relying on the C++14 standard and should be compilable on every platform, that is supported by Qt and Eigen. The model implementation makes use of object-oriented programming to easily implement new models. It is out of the scope of this article to deal with the detailed implementation, but a short summary will be given:

The source code is separated into four parts: (1) the core components containing the models, source code for optimisation and collected mathematical tools. Statistical analysis is implemented in the second part (2). Both parts, (1) and (2), are independent of any user interface and provide the functionalities for the pure command line application *suprafit_cli.exe* (3) and the graphical user interface *suprafit.exe* (4).

The core part holds the functionality to store the experimental data (*DataClass*), that is realised using a shared data pointer. Model preparation is done in the abstract class *AbstractModel*, that is based on that *DataClass*. Therefore, each implemented model inherits from *AbstractModel* and *DataClass*, respectively (Fig. 2). In the specific model implementation, the equations of the model and the number of input parameters have to be defined, as well as the names of each parameter. More details can be found in the source code documentation for the *AbstractModel*, *AbstractTitrationModel* and *Michaelis-Menten-Model*.[37]

Parallelisation is mostly done using the threads concept utilising QThreadPool and QRunnable, but individual parts use openMP. Data storage is done using the JSON Format (**.json*) or Zip compressed JSON (**.suprafit*).

4.3 Statistical tools and further analysis

4.3.1 Confidence Intervals

Parameter (θ) estimation is the main question in regression, as it allows the rational analysis and comparison of data sets and experiments. Yet the knowledge of θ is often not sufficient for rational analysis,[38] as the best fit values may differ for several performed experiments. The confidence interval of a parameter θ_i estimates the range $[\theta_{i,-}, \theta_{i,+}]$, within which the true parameter $\tilde{\theta}_i$ can be expected. However, the standard approach used in (multiple) linear regression cannot be applied for non-linear problems. *SupraFit* provides two basic routes to approximate the confidence interval, both being described in the literature before. Explicit references will be given in each section. One of the goals of this article and *SupraFit* regarding the statistical tools is not to have one correct way to calculate confidence intervals, but rather present the already known techniques, provide an easy way to access those and show some examples on how these methods can be applied to parameter estimation problems.

4.3.2 Confidence Intervals by Monte Carlo simulations and percentile method

A powerful tool, that is used in many fields of science is the Monte Carlo simulation.[39] It has already been applied to both

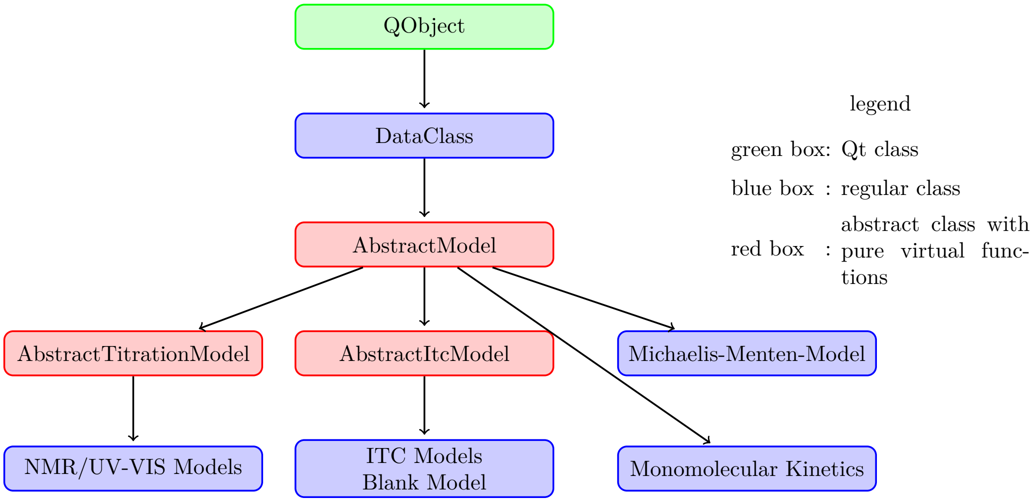


Fig. 2 Inheritance relationship in *SupraFits* model implementation. To implement a new model, a C++ *class* has to be derived from *AbstractModel class* and the most important virtual functions have to be implemented.

ITC and NMR titration apart from confidence calculation.[40–42] The application to calculate confidence intervals has been reported for titration experiments by Thordarson[15] and in general by Motulsky and Christopoulos.[43] The confidence intervals from Monte Carlo simulations are obtained using the percentile method, which has been discussed alongside with resampling methods by Efron.[38] Efron noted, that the section dealing with confidence intervals ”is highly speculative in content.”[38]

The basic idea of the Monte Carlo approach is to theoretically repeat the performed experiment several times (T). A single theoretic step is being realised by adding a random error ε_i to $y_{calc,i}$ and then obtain a new set of data mimicking the original experimental data including realistic errors. These data can be used to estimate a new set $\underline{\theta}$. Performing these steps T times is denoted as Monte Carlo (MC) simulation within this context.

Two main approaches to define the errors ε are implemented in *SupraFit*, (a) they are calculated from the standard normal distribution $\varepsilon \in N(\mu = 0, \sigma^2)$ or (b) randomly chosen from the absolute errors obtained after the successful fit ($\varepsilon \in \underline{e}$). The later will be called bootstrapping (BS) in *SupraFit* and may be interpreted as a mixture of a typical Monte Carlo simulation and resampling technique. Bootstrapping is one of the resampling plans discussed by Efron.[38,44] More recent discussions and problems using the bootstrapping method can be found in Canty et al.[45] and in Efron and Hastie.[46]

The applied standard deviation σ_{MC} in approach (a) can be taken from the SE_y, σ_{fit} or as manually defined value, where SE_y is the default choice as proposed by Motulsky and Christopoulos since it is the corrected standard deviation (eq. 31). The $1-2\alpha$ confidence interval for each model parameter is then calculated using the percentile method:

$$\theta_{i,-} = C\hat{D}F^{-1}(\alpha) \qquad \theta_{i,+} = C\hat{D}F^{-1}(1-\alpha) \tag{33}$$

which results in the 95% confidence interval if $\alpha = 0.025$. In *SupraFit*, this is realised by collecting all model parameters for each Monte Carlo step and then take $\alpha \cdot T$ and $(1-\alpha) \cdot T$ entry of the ordered list of the corresponding parameter. More advanced percentile methods, which are available in octave or R, are not implemented, so for a smaller number of T the results differ from those obtained with the standard approach using the quantile function in octave or R.[¶] More robust methods will be implemented in future releases. Efron proposed 2000 steps as minimum for bootstrapping methods,[46] which is taken as standard for all Monte Carlo simulations in conjunction with the percentile method. Since Monte Carlo simulations are parallelised,[||] it benefits from the multicore architecture of modern desktop computers. Monte Carlo results are then reported as histogram-like charts as printed in Fig. 3. The box represents the 95% confidence interval, the dash-dotted line the estimated parameter. The individual bins are not plotted as typical bars but rather as a line-plot.

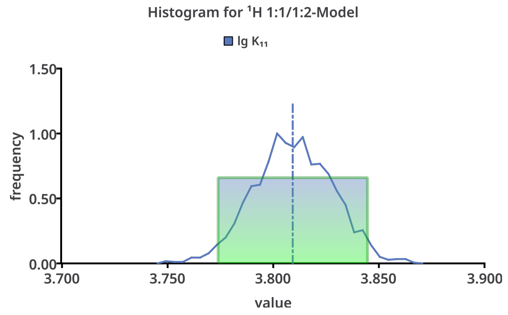


Fig. 3 Standard representation of a histogram-like chart obtained after performing a Monte Carlo simulation.

Alternatively to the variation of y_{calc}, Thordarson proposed the variation of input data, which are the initial concentrations of host and guest molecules in case of NMR titration.[15] This derivation can be performed alongside with standard Monte Carlo simu-

[¶] See https://octave.org/doc/v4.0.1/Descriptive-Statistics.html. Last visit 17.01.2022.

[||] Monte Carlo simulation are spawned across the threads, that roughly each thread performs $T/NThreads$ optimisation.

lations. To the best of the authors knowledge confidence interval calculations have not been reported for this derivation, however percentiles can be calculated in the same way.

4.3.3 Confidence Intervals using the F-Test approach

The F-Test approach to confidence intervals has first been proposed by Box[47] and Beale,[48] and further outlined by Beechem[49] as well as Bates and Watts.[50] Taking the least-squares estimated set of parameters $\hat{\theta}$, the confidence interval then includes all values θ that are equal to the best-fit estimation $\hat{\theta}$. This can be formulated as following hypotheses $H_0 : \theta = \hat{\theta}$ and the alternative $H_A : \theta \neq \hat{\theta}$. The decision is based on the F-Test (eq. 34), where the ratio of $SSE(\theta)$ and $SSE(\hat{\theta})$ has to be smaller than the value, that defines the $(1-\alpha)\cdot 100\%$ confidence interval.[51]

$$\frac{SSE(\theta) - SSE(\hat{\theta})}{SSE(\hat{\theta})} \leq \frac{K}{N-K} F^{\alpha}_{N,N-K} \tag{34}$$

$$SSE_{max} = SSE(\theta) \leq SSE(\hat{\theta}) \cdot \left(1 + F^{\alpha}_{N,N-K} \frac{K}{N-K}\right) \tag{35}$$

In equation 34, K refers to the number of parameters, N to the number of data points and $F^{\alpha}_{N,N-K}$ to the critical value in the F-distribution for the given degrees of freedom and desired confidence interval. A graphical interpretation is given in Fig. 4. The sum of squares has a minimum at $SSE(\hat{\theta})$ and can be decreased to $\theta_{i,-}$ or increased to $\theta_{i,+}$ while the error is smaller than SSE_{max}.

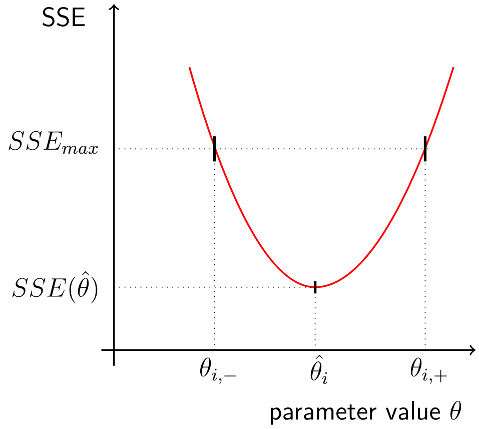


Fig. 4 Graphical interpretation of the F-Test approach. The confidence interval $[\theta_{i,-}, \theta_{i,+}]$ is not necessarily symmetric.

At least two different approaches to the F-Test are mentioned in the literature (a) the Weakened Grid Search[15,49] (WGS) and (b) Model Comparison (MOC).[11,43] Keller at al.[52] published an Excel-Guide to apply the F-Test to Michaelis-Menten Kinetics using the Weakened Grid Search. *SupraFit* provides both approaches to the F-Test, that will now be introduced:

Weakened Grid Search

Having K parameters to be analysed, the first θ_i is changed by small δ_θ[**] and then fixed, while the remaining $\theta_{j\neq i}$ are optimised. The parameter θ_i is changed again by δ_θ and the $\theta_{i\neq j}$ parameters are estimated anew. This is to be repeated as long as $SSE(\theta)$ is smaller than SSE_{max} and therefore H_0 is not rejected. This procedure is performed for all parameters in the same manner and all θ that satisfy equation 35 define the confidence region.[15] In *SupraFit* some additional parameters are introduced to control the procedure, like the maximum number of steps, the step size and the convergence threshold for the sum of squares. A comprehensive list is given in the manual of *SupraFit* and a short description of each parameter is shown as tooltip in the *SupraFit* program. Obtained results are graphically presented as shown in Fig. 5, where one parameter was analysed. The dash-dotted line indicates the estimated value $\hat{\theta}_i$ and the solid line indicates the obtained sum of squares for each variation of θ_i while $\theta_{i\neq j}$ are being optimised. Only values where the error is smaller than the threshold are plotted. The Grid Search is parallelised, so that for each parameter θ_i two processes independently evaluate either $\theta_{i,-}$ or $\theta_{i,+}$.

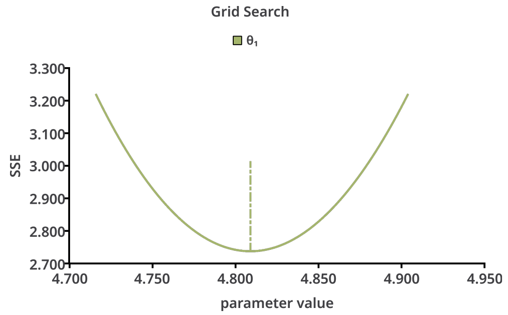


Fig. 5 Sample representation of the Weakened Grid Search result.

Model Comparison

An alternative way to the F-Test approach is denoted as Model Comparison. During MOC calculations θ_i is varied by an amount of δ_θ while the remaining $\theta_{i\neq j}$ are not optimised, but systematically varied to fullfill equation 35. The parameter θ_i is then again changed by δ_θ and the remaining $\theta_{i\neq j}$ are varied to meet the condition in equation 35. This is repeated until the change of θ_i disobeys equation 35. After performing this approach for all K parameters, the limits of the confidence region can be extracted from all obtained values of θ_i as $\theta_{i,-} = min(\theta_i)$ and $\theta_{i,+} = max(\theta_i)$.[43] Assuming that there is only one parameter to be optimised, applying WGS and MOC as described in Motulsky and Christopoulos,[43] both methods perform similarly: θ_k will be varied by $\pm\delta_\theta$ until $SSE(\theta_k)$ reaches the maximum possible SSE and the tuple $(\theta_{k,+}, \theta_{k,-})$ correspondence to the confidence interval. This approach of continuously varying one parameter is implemented in *SupraFit* as Simplified Model Comparison

[**] δ_θ is both, positive and negative so that θ_i is tested for values smaller and greater than $\hat{\theta}_i$.

(SMOC). Like WGS, the Simplified Model Comparison benefits from multiple processes, since each parameter is evaluated in a single thread.

Instead of systematic variation, *SupraFit* provides the Model Comparison as Monte Carlo experiment just like the calculation of an arbitrary area: Uniform random numbers are generated within defined boundaries for every θ_i, where these random parameters are stored if $SSE(\theta)$ meets equation 35. The confidence interval is then defined by the minimum and maximum values for all θ. The implementation works as follows: Simplified Model Comparison is applied to each parameter and the confidence interval $[\theta^{SMOC}_{i,-}, \theta^{SMOC}_{i,+}]$ is obtained. The intervals are scaled by variable parameters, which define a rectangular box in case of two variables, a cuboid for three parameters etc. (dash-dotted box in Fig. 6). Uniform random numbers are generated within the interval defined by the box and checked if they obey equation 35. If they do, the parameters are kept, otherwise they are discarded. An ideal confidence interval is represented in Fig. 6 as red ellipsoid, with the maximum values for θ_1 and θ_2 form the limits of the confidence interval. Similar to previous methods, Model Comparison is parallelised, where amount of Monte Carlo steps is equally divided across the threads.

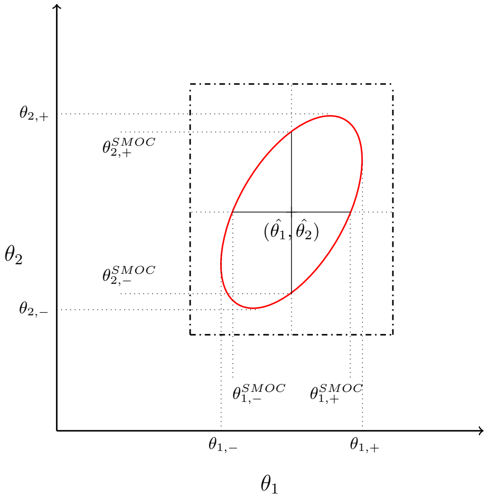


Fig. 6 Calculation of the confidence interval using the Model Comparison and the Monte Carlo approach. Random values of θ_1 and θ_2 are generated within the dash-dotted boundaries. If $SSE(\theta_1, \theta_2)$ meet equation 35, the parameters are kept.

4.3.4 Resampling Methods

Cross Validation (CV) is a powerful tool, applied for example in QSAR in conjunction with principal component selection.[53] In *SupraFit*, CV will be applied to determine the sufficiency of the used model. Another method, not yet described and applied to supramolecular titration experiments is called "Reduction Analysis." Both methods will be introduced in a subsequent article, that focuses on a statistical approach to analyse binding stoichiometry.

4.4 Linear Regression Tool

SupraFit provides a linear regression tool for experimental data, that can be used to fit several linear functions to experimental data. The data points are continuously divided: In case of three functions, the first function is fitted using the first $n_1 ... n_i$ data points, the next functions uses the next $n_{i+1} ... n_j$ data points and the last functions uses the remaining $n_{j+1} ... n_N$ data points. The maximal number of functions is $N/2$, where each function is described by two points. The currently implemented method tests all available combinations and returns an ordered list. One field of use will be shown for NMR titration, to create Mole Ratio plots. Another application will be shown within the ITC examples.

4.5 Global fitting

Programs like *pytc* or *SEDFIT* provide methods to perform a global fit,[16,21] that is to fit a single set of parameters to more than one experiment. In that fashion, analysing several signals in NMR titrations is already a global fit,[15] since one formation constant is connected to two or more signals. While a global fit for NMR titration is straightforward, combining several ITC experiments is performed with *MetaModels* in *SupraFit*. *MetaModels* are empty container models, that can hold and manage real models. Model parameters can be handled individually or any in combination thereof. However, the first approach is identical to a local fit. Statistical analysis or global search can be performed on *MetaModels* in the same way as on simple models. An example of *MetaModels* will be discussed in the ITC section.

5 Examples

5.1 Model function with uncorrelated and correlated parameters

Uncorrelated parameters

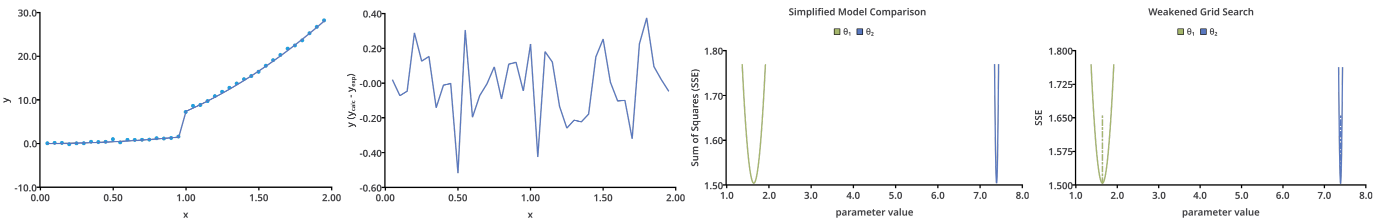
An example using a function with two uncorrelated parameters θ_1 and θ_2 is used to illustrate the preceding aspects of the statistical analysis. The function in equation 36 acts on the element m of the vector having M elements:

$$f(\theta_1, \theta_2) = \begin{cases} \theta_1 x^2 & \forall m < M/2 \\ \theta_2 x^2 & \forall m \geq M/2 \end{cases} \tag{36}$$

Thus, θ_1 acts on the first half of the interval while θ_2 acts on the second half. Depending on the values for θ, the function is discontinuous at $m = M/2$. In the range of $\{0.05, 0.05+0.05, ..., 1.95-0.05, 1.95\}$ with $\theta_1 = 1.8801$ and $\theta_2 = 7.4043$, after adding a random error ($\varepsilon \in N(0, 0.25)$), the function (eq. 36) is drawn in Fig. 7.

The 95% confidence intervals using F-Test based methods applying the (Simplified) Model Comparison and Weakened Grid Search approach are given in Table 6. Both have been applied to either parameters individually (MOC^a, WGS^a) or to both (MOC^b, WGS$^{b)}$) together. The F-Test confidence intervals are effectively the same, independent of the approach, with some numerical differences due to step size during the evaluation. Using Monte Carlo simulation ($\varepsilon = SEy$) with the percentile method, the confidence interval is much narrower than these obtained with the F-Test approach. Those differences were already pointed out by Motulsky and Christopoulos.[43]

The variation of the individual parameters θ_i by $\pm\delta_{\theta,i}$ and the



(a) (b) (c) (d)

Fig. 7 Representation of (a) a sample function with two uncorrelated parameters and (b) the added normal distributed error as well as the variation of θ and the corresponding SSE during the (c) Simplified Model Comparison and (d) Weakened Grid Search.

Table 6 95% Confidence Intervals obtained after Simplified Model Comparison (SMOC), Weakened Grid Search (WGS), Model Comparison (MOC) and Monte Carlo simulation (MC). [a]Both parameters are analysed individually. [b]Both parameters are analysed at the same time.

	$[\theta_{1,-}, \theta_{1,+}]$	$[\theta_{2,-}, \theta_{2,+}]$
SMOC	1.3674 - 1.9157	7.3429 - 7.4381
MOC[a]	1.3680 - 1.9151	7.3429 - 7.4381
MOC[b]	1.3680 - 1.9151	7.3430 - 7.4380
WGS[a]	1.3676 - 1.9156	7.3435 - 7.4375
WGS[b]	1.3675 - 1.9157	7.3429 - 7.4381
MC	1.4217 - 1.8433	7.3543 - 7.4230

corresponding SSE for SMOC and WGS are shown in Fig. 7(c) and 7(d). In both charts, the series show a parabolic trend, indicating that the SSE_{max} can be reached during variation.

The correlation coefficient for θ_1 and θ_2 and the scatter plots (Fig. 8) after MOC, WGS and MC clearly indicate that there is no dependency between both parameters, which is in agreement with the given function. The obtained correlation coefficient for θ_1 and θ_2 is $3.6 \cdot 10^{-5}$ after Model Comparison. Using WGS the accepted values for θ_1 and θ_2 show a correlation coefficient of zero. The lines display two sets of accepted values for θ_1 and θ_2, where one parameter is not affected by changing the other. The model parameters after Monte Carlo simulation indicate no correlation ($R^2 = 1.9 \cdot 10^{-4}$) as well, but the pairs of θ_1 and θ_2 do not form a complete ellipse as obtained after Model Comparison. However, in case of functions or models with uncorrelated parameters the implemented F-test based approaches lead to practically identical results, which differ from the Monte Carlo simulation based results.

Correlated parameters

A function where θ_1 and θ_2 are not independent is given in equation 37. The same input data are used as in previous example, where $\theta_1 = 4.8321$ and $\theta_2 = 8.5912$. Random error ($\varepsilon \in N(0, 0.25)$) is added to simulate experimental noise.

$$f(\theta_1, \theta_2) = ld(\theta_1) \cdot ld(\theta_2) \cdot x^2 + \frac{ld(\theta_1)}{x} + ld(\theta_2) \cdot x \tag{37}$$

The corresponding diagrams are plotted in Fig. 9, including the graphical interpretation of the SMOC and WGS approaches.

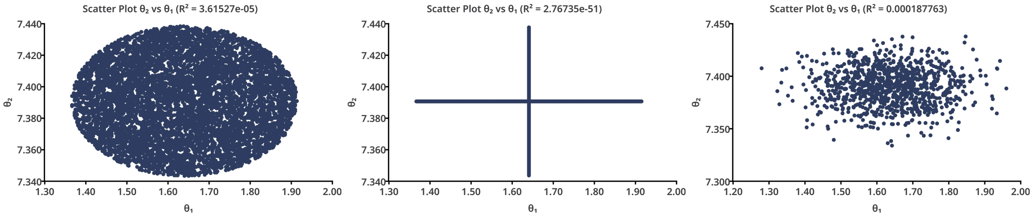
The confidence intervals, that are calculated similarly to the previous example, are summarised in Table 7: SMOC and MOC[a] result in the same confidence interval, and MOC[b] and both WGS[a] and WGS[b] result in the same confidence intervals, however different from the first one. This is expected, since SMOC and MOC[a] take only one parameter into account and fix the remaining, while MOC[b] and WGS take both parameters into account. The confidence intervals after Monte Carlo simulation are narrower than the WGS/MOC[b] confidence intervals, but broader than the SMOC and MOC[a] intervals.

Table 7 95% Confidence intervals obtained after Simplified Model Comparison (SMOC), Weakened Grid Search (WGS), Model Comparison (MOC) and Monte Carlo simulation (MC). [a]Both parameters are analysed individually. [b]Both parameters are analysed at the same time.

	$[\theta_{1,-}, \theta_{1,+}]$	$[\theta_{2,-}, \theta_{2,+}]$
SMOC	4.7583 - 4.8601	8.5829 - 8.8463
MOC[a]	4.7583 - 4.8601	8.5830 - 8.8463
MOC[b]	4.7164 - 4.9037	8.4773 - 8.9618
WGS[a]	4.7160 - 4.9038	8.4760 - 8.9620
WGS[b]	4.7160 - 4.9030	8.4766 - 8.9616
MC	4.7381 - 4.8816	8.5294 - 8.9031

The graphical interpretation of SMOC and WGS are shown in Fig. 9(c) and 9(d). While all series again show a parabolic trend, the series for θ_1 or θ_2 differ slightly for both methods. The correlation between θ_1 and θ_2 can be analysed using the correlation coefficient and the scatter plots as shown in Fig. 10. Apart from the different confidence intervals, the ellipsoid after MOC is rotated with respect to the ellipsoid in Fig. 8a and correlation can be observed ($R^2 = 0.70$). The scatter plot after WGS shows two lines again, where each line is assigned to the variation of one parameter. The correlation coefficient indicates a strong correlation ($R^2 = 0.98$), which however is an artefact since only the best-fit values are included but not all possible values that obey equation 35. Monte Carlo simulation on the other hand leads to a similar scattering of the parameters and a very similar correlation coefficient ($R^2 = 0.70$).

Having two parameters (θ_k and θ_l) and performing WGS for only one parameter θ_k, the F-Test confidence interval for the corresponding parameter is obtained since θ_j is always adjusted. However, performing the MOC and limiting it to one parameter θ_k, the confidence interval will always be smaller or equal to the correct F-Test confidence interval, since at $SSE(\theta_k^{MOC}) = SSE_{max}$ there is still the other parameter θ_l to be adjusted. If there is no

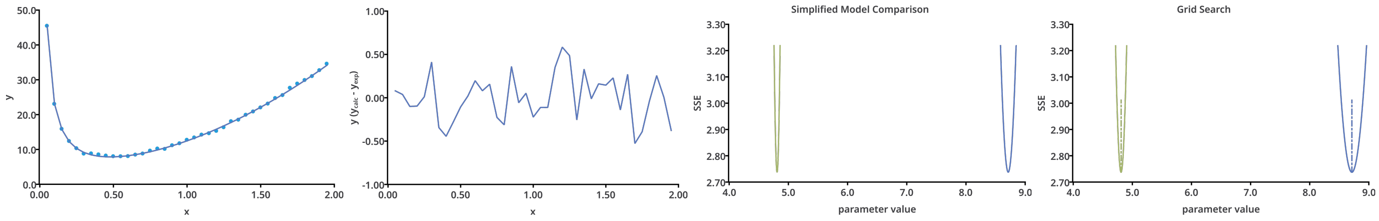


(a) MOC, $R^2 = 3.62 \cdot 10^{-5}$

(b) WGS, $R^2 = 2.77 \cdot 10^{-51}$

(c) MC, $R^2 = 1.88 \cdot 10^{-4}$

Fig. 8 Scatter plots after confidence calculation using (a) Model Comparison, (b) Weakened Grid Search and (c) Monte Carlo simulation for the model with uncorrelated parameters.



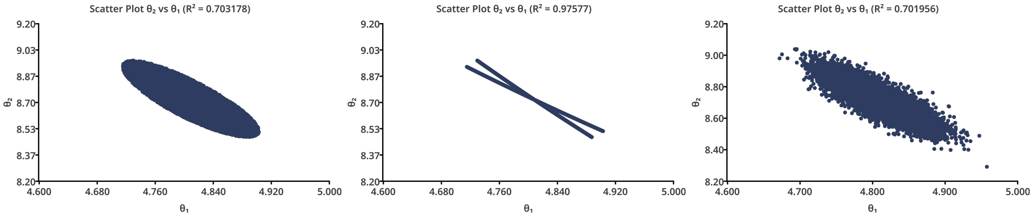
(a)

(b)

(c)

(d)

Fig. 9 Representation of (a) a sample function with two correlated parameters and (b) the added normal distributed error as well as the variation of θ and the corresponding SSE during the (c) Simplified Model Comparison and (d) Weakened Grid Search.



(a) MOC, $R^2 = 0.70$

(b) WGS, $R^2 = 0.98$

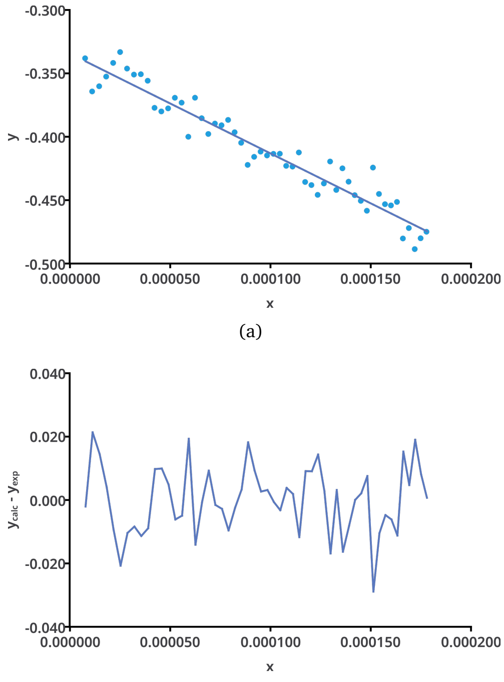
(c) MC, $R^2 = 0.70$

Fig. 10 Scatter plots after confidence calculation using (a) Model Comparison, (b) Weakened Grid Search and (c) Monte Carlo simulation for the model with correlated parameters.

correlation between θ_k and θ_l, both parameters can be varied independently of each other and the F-Test confidence interval of the Simplified Model Comparison and WGS are equal.

5.2 Linear Regression

In case of linear models, the (1-α) confidence intervals can be calculated with standard software like Excel or similar spreadsheet programs as well as statistical software like R. In Table 8, we report the confidence intervals for a linear model using the t-distribution approach calculated using *Gnumeric*[54] as well as the approaches for non-linear models implemented in *SupraFit*. The data used were obtained adding $\varepsilon \in N(\mu = 0, 0.01)$ to a linear model $y = \theta_1 x + \theta_2$ with $\theta_1 = -820.000$ and $\theta_2 = -0.333$ (Fig. 11). The least-squares estimated parameters are $\theta_1 = -787.551$ and $\theta_2 = -0.334$.



(a)

(b)

Fig. 11 Representation of (a) a linear function and (b) the added normal distributed error.

The non-linear F-Test based confidence interval differs much from the smaller linear t-distribution bases interval. Monte Carlo simulation with $T = 50000$ steps was performed as bootstrapping and using SE_y and σ_{fit} as input standard deviation. The BS confidence interval is the smallest and the interval using SE_y is the widest, since $SE_y > \sigma_{fit}$. However, the obtained confidence intervals after Monte Carlo simulations are very close to the one calculated with the linear approach, being only slightly smaller. Using SE_y as ε for Monte Carlo simulation recovers the linear approach best.

Table 8 95% Confidence Intervals obtained after Linear Regression, Weakened Grid Search and Monte Carlo simulation (T = 50000).

	$[\theta_{1,-}, \theta_{1,+}]$	$[\theta_{2,-}, \theta_{2,+}]$
linear	[-846.7845 , -728.3161]	[-0.3408 , -0.3280]
WGS	[-861.9370 , -713.1640]	[-0.3424 , -0.3264]
	Monte Carlo simulations	
SE_y	[-845.0710 , -729.5970]	[-0.3406 , -0.3282]
σ	[-844.3855 , -730.3190]	[-0.3406 , -0.3282]
BS	[-843.3700 , -732.0910]	[-0.3404 , -0.3283]

5.3 NMR Titration

To demonstrate the application of *SupraFit* in case of NMR titration, example calculation on an artificial NMR titration with a 1:1/1:2 binding stoichiometry were performed. The stability constants to set up the experimental data were chosen to be $lgK_{11} = 3.81$ and $lgK_{12} = 2.14$ The chemical shifts can be found in the supporting information. The individual shifts are not meant to represent a realistic example. A random error obtained from a normal distribution with $\varepsilon \in N(\mu = 0, 0.001)$ was added afterwards, where every single signal has the same σ, therefore e.g. signal 6 ($\Delta\delta = 2.3038\ ppm$) and signal 7 ($\Delta\delta = 0.2441\ ppm$) have both the same random error. The "experimental" titration curve can be found in Fig. 12(a). The four possible models (1:1, 2:1/1:1, 1:1/1:2 and 2:1/1:1/1:2) were tested without cooperative relationships.

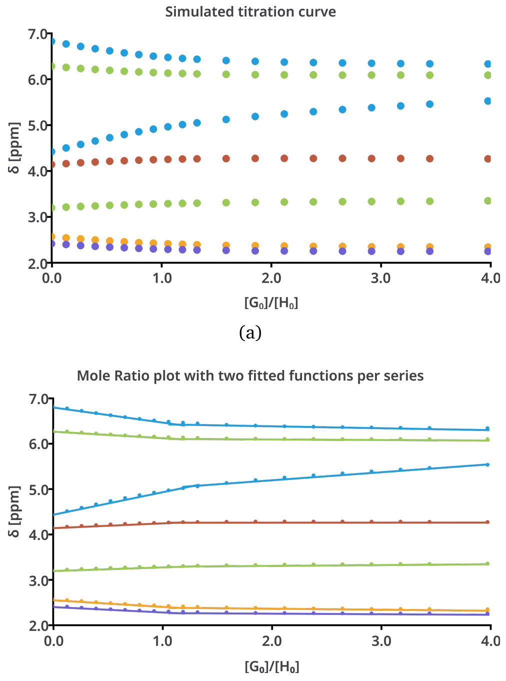
Mole Ratio Plot

Using *SupraFits* linear regression method with two functions, Mole Ratio plots can easily be generated. Since the typical plots exhibit the chemical shift on the x axis and the ratio on the y axis, the plots obtained using *SupraFit* differ from the "standard" plots. However, as the chemical shifts depends on the ratio and not vice versa, *SupraFit* provides only the "non-standard" way. The plot can be found in Fig. 12(b). For each series, all possible intersections of adjacent linear functions are calculated. The result for the best fit, that fit minimising the sum over all *SSE*, is listed in the supporting information. The intersections of the two functions per signal ranges between 1.13 and 1.27, indicating a system that exhibits 1:2 species. This is in accordance with the stoichiometry of the original model.

Fitted parameter

The resulting stability constants (lg K) after optimisation are printed in Table 9, statistical judgements using *SSE* and SE_y can be found in Table 10. The titration curve as well as the remaining absolute errors can be found in Fig. 13. The complex formation constants for the correct model differ only slightly from the initial ones. The easier 1:1 model estimates a lgK_{11} that is too small, as happens upon fitting the 2:1/1:1 model. The most complex model resamples lgK_{11} and lgK_{12}, but the incorrect model parameter lgK_{21} is realistic. Some of the chemical shifts in the 2:1/1:1 are smaller than zero ($\delta_{A_2B,1} = -6.6334\ ppm$), indicating a change in the chemical shift up to 13 ppm ($\delta_{AB,1} = 6.5565\ ppm$). A full list of all parameters can be found in the example file in the *SupraFit* repository at GitHub.

The "visual inspection" as described by Hynes,[9] can be performed using the charts in Fig. 13(c) and 13(d), where all absolute errors are plotted in Fig. 13(c) and the errors only from the 1:1/1:2 model and 2:1/1:1/1:2 model are plotted in Fig.



(a)

(b)

Fig. 12 (a) Simulated titration curves with $lgK_{11} = 3.81$ and $lgK_{12} = 2.14$ and seven observed signals. (b) The Mole Ratio plots show an intersection of two linear functions at a molar ration between 1.0 and 1.5.

Table 9 Estimated lg K values for the applied 1:1, 2:1/1:1, 1:1/1:2 and 2:1/1:1/1:2 models

model	lgK_{21}	lgK_{11}	lgK_{12}
true model		3.8100	2.1400
1:1		3.0991	
2:1/1:1	1.7448	2.6694	
1:1/1:2		3.8092	2.1090
2:1/1:1/1:2	1.9893	3.8063	2.0429

13(d). Clearly the 1:1 model perform worst, followed by the 2:1/1:1 model, with both having heteroscedastic errors. The remaining two models are optically indistinguishable with both errors being homoscedastic.[††] Considering the resulting SSE, the decision towards the correct model can already be made, since $SSE_{1:1/1:2} \approx SSE_{2:1/1:1/1:2}$ and $3 \cdot SSE_{1:1/1:2} < SSE_{1:1}$.[15] Comparing SSE of the fitted 1:1/1:2 model and the correct model show the slightly smaller error for the optimised model.

Table 10 The sum of squared errors (SSE) as well as σ and SE_y after testing four models on the simulated data set. [a]Not calculated, since this model is not fitted to the data

model	parameter fitted	SSE	SE_y	σ
1:1	15	0.036459	0.017078	0.016196
2:1/1:1	23	0.001761	0.003878	0.003560
1:1/1:2	23	0.000132	0.001062	0.000975
2:1/1:1/1:2	31	0.000127	0.001077	0.000954
fitted 1:1/1:2	23	0.000132	0.000983	0.000975
correct 1:1/1:2	-	0.000165	-[a]	0.001088

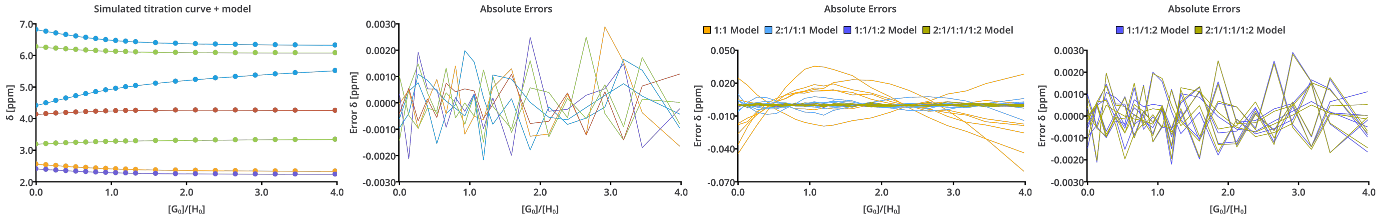
5.3.1 Monte Carlo Confidence Intervals

Following the strategy of the Monte Carlo simulation, the introduced error can be calculated from the standard normal distribution with (a) a defined variance or (b) via bootstrapping. To test the influence of different approaches on the confidence interval, a set of simulations were performed on the given dataset with the optimised 1:1/1:2 model. The standard normal distributed errors were generated with $\sigma_{MC} = SE_y$, $\sigma_{MC} = \sigma_{Fit}$, $\sigma_{MC} = 1e^{-3}$, $\sigma_{MC} = 2e^{-3}$, $\sigma_{MC} = 3e^{-3}$ and $\sigma_{MC} = 5e^{-3}$. Monte Carlo simulation with T = 100, 200, 300, 500, 700, 1000, 1500, 2000, 2500, 3000 and 5000 steps were performed, where each simulation was repeated 300 times. The 95% confidence interval was then characterised by the median and standard deviation of the 0.95 interpercentile ranges (IPR) for these 300 Monte Carlo simulations.

The boxplots and the standard deviation of the 0.95 IPR values for the stability constants lgK_{11} and lgK_{12} after the Monte Carlo simulation are reported in Fig. 14 and show expected behaviour: With increasing steps T, the observed standard deviation of the IPR decreases. The same trend is visible for the other Monte Carlo simulations including BS (see Fig. S7 - S13). With increasing step count the IPR converges to the ideal IPR that could be obtained after an infinite number of steps. As Efron stated,[46] at least 2000 steps are required for the bootstrap method to obtain reliable results. However, since every Monte Carlo step requires the least-squares estimation of θ, this approach is demanding. As shown in Fig. 15, Monte Carlo simulation scales well with the number of threads used and benefits from Hyperthreading technology.[‡‡] Therefore, accurate Monte Carlo simulation with 2000

[††] This is expected as they resample the original normal distributed random numbers.

[‡‡] The benchmark was obtained on a i9-7920X CPU with 12 cores overlocked to 4.00GHz, using openSUSE 15.0 Leap. *SupraFit* was compiled using gcc 7.4.1.



(a) (b) (c) (d)

Fig. 13 (a) Chemical shifts and fitted curves using an 1:1/1:2 model ($lgK_{11} = 3.81$ and $lgK_{12} = 2.11$) and (b) the resulting absolute errors. (c) The absolute errors for all four models are plotted in one chart, showing that the 1:1 model and 2:1/1:1 performe worse than the 1:1/1:2 and the 2:1/1:1/1:2 model. (d) Both models, 1:1/1:2 and 2:1/1:1/1:2, show similar residuals.

steps can easily be obtained within minutes even on a desktop computer with fewer cores.

As shown in Fig. 16 the confidence intervals obtained from 300 Monte Carlo simulations with each simulation performed with 5000 steps using BS or random errors and different σ_{MC} differ. As σ_{MC} increases, the confidence interval gets broader and standard deviation of the IPR increases. However, the differences between bootstrapping and random error with $\sigma_{MC} = \sigma_{fit}$ are very small but since the Kruskal-Wallis-test results in a p-value = 0.002 < 0.05 for lgK_{11} and p = 0.023 < 0.05 for lgK_{12}, the differences are significant for the given example. The corresponding plots for lgK_{12} are presented in Fig. S14.

5.3.2 Correlation of lgK_{11} and lgK_{12}

Since the current NMR titration model has more than two parameters, the correlation of lgK_{11} and lgK_{12} will be analysed either neglecting or taking the parameters, e.g. the chemical shifts, into account. Therefore, a Monte Carlo simulation with $\sigma_{MC} = SE_y$ and T = 10000, two runs of Weakened Grid Search, the first only for lgK_{11} and lgK_{12} and the second for all parameters and Model Comparison for lgK_{11} and lgK_{12} were performed. The scatter plots for lgK_{11} vs lgK_{12} are shown in Fig. 17 and the confidence intervals are given in Table 11.

Table 11 95% Confidence Intervals obtained after Weakened Grid Search (WGS), Model Comparison (MOC) and Monte Carlo simulation (MC). [a]Only lgK_{11} and lgK_{12} where analysed and [b]all parameter were analysed.

	$[lgK_{11,-}, lgK_{11,+}]$	$[lgK_{12,-}, lgK_{12,+}]$
WGS[a]	3.698 - 3.926	1.935 - 2.242
WGS[b]	3.697 - 3.927	1.934 - 2.243
MC	3.773 - 3.846	2.059 - 2.155
MOC	3.801 - 3.818	2.104 - 2.114

The first two charts show the scattering of the complex formation constants after applying the Weakened Grid Search, where Fig. 17(a) contains only two series, since only two parameters were tested. However, the chart in Fig. 17(b) shows more than two series, as all parameters were taken into account. Incorporating more parameters, the correlation coefficient drops from 0.80 to 0.74 since more points from the original series are available. However, the high correlation is an artefact as already pointed out in the example of the function with correlated parameters in the previous section. The scatter plot after Monte Carlo simulation in Fig. 17(c) shows an ellipsoid, with the parameters having a correlation coefficient of 0.37. On the other hand, using Model Comparison with only taking two parameters into account, one obtains a complete ellipsoid, which however is rotated with respected to the Monte Carlo ellipsoid and to the series obtained after Weakened Grid Search (Fig. 17(d)). Therefore, naive Model Comparison leads to wrong results regarding confidence intervals and the ellipsoid, if correlated parameters are ignored.

5.4 Isothermal titration calorimetry

The ITC data used in the following section are taken from the *pytc-demo*. The complex formation of Calcium with EDTA (see https://github.com/harmslab/pytc-demos, last checked 17.01.2022) where reported by Harms et al. [16] to demonstrate the *pytc* tool. The heat is given in cal and cal/mol. In the first part, the initial guess of the parameters in case of a 1:1 model are described, since a good starting point for the non-linear regression is essential.

The fx value, the inflection point of the titration curve, [55] is guessed by fitting three non-overlapping linear functions to the isotherm. The guessed fx value is then obtained as mean of the intersection of first with the second function and the second with the third function (Fig. 18). The heat of formation is calculated using the heat of the third injection $Q_{2,3}$ divided by the change in concentration of the added guest $[B]$ component. It is assumed, that at the start of the titration the concentration of the formed complex is nearly the same as the added guest concentration since $[B] \ll [A]$. The stability constant is then calculated using the bisection method within the limits of $1 \leq lgK_{11} \leq 10$. The initial guessed parameters of the 1:1 model are applied to the models of mixed stoichiometries as well. See Table S2 for the comparison of the initial guessed and fitted parameters for the *hepes* data.

Global Fit

MetaModels were used to globally fit lgK_{11} and δH_{AB} to the data of *hepes-01*, *hepes-02* and *hepes-03* from the *pytc-demo* that are followed by Monte Carlo simulation to estimate the confidence intervals. These results were then compared to the confidence intervals obtained from Monte Carlo simulations for the individual experiments. The obtained parameters and the confidence intervals using Monte Carlo simulation ($\sigma_{MC} = SE_y$, 5000 steps) are listed in Table 12. While the globally estimated lgK_{11} is nearly the mean of the individual models (7.595), the IPR for lgK_{11} can not be approximated by the mean of the individual IPR (0.065). The

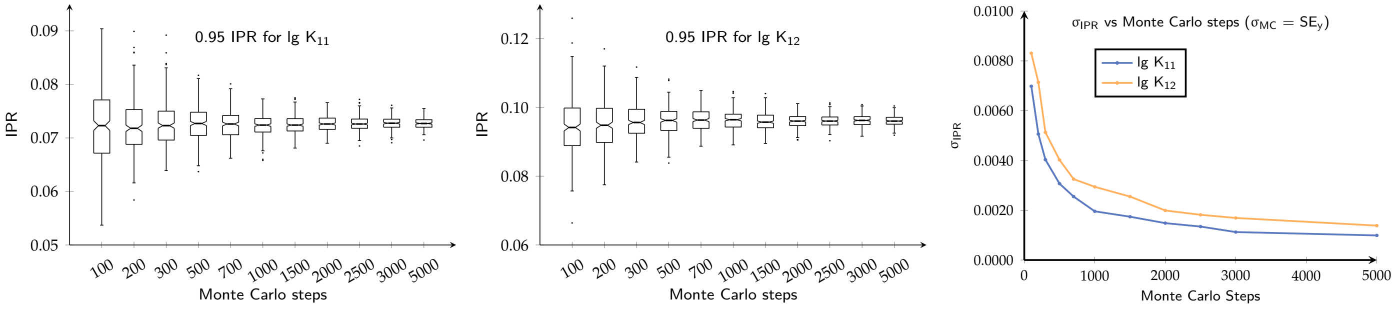


Fig. 14 Variation and standard deviation of the IPR for lgK_{11} and lgK_{12} after several Monte Carlo simulations with $\sigma_{MC} = SE_y = 1.062 \cdot 10^{-3}$.

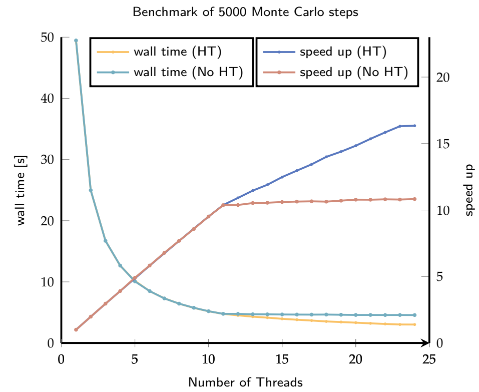


Fig. 15 Wall time in seconds and speed up as function of the number of threads used in Monte Carlo simulation. The wall time is averaged over 50 runs. The benchmark was performed on a Intel i9-7920X CPU @ 4.00GHz (12 physical cores, overclocked) with and without Hyperthreading (HT).

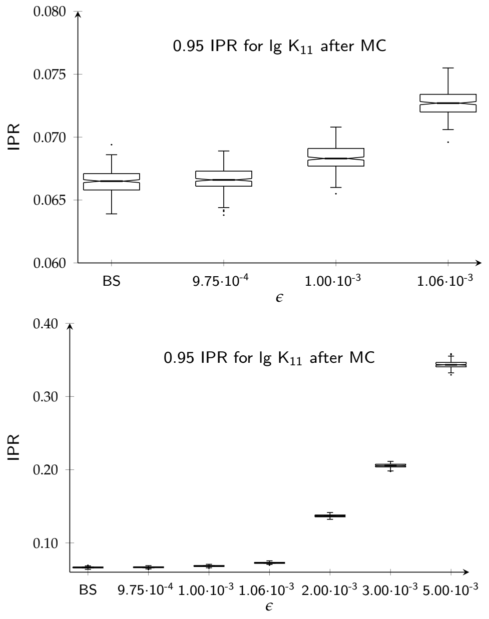


Fig. 16 IPR of lgK_{11} for several Monte Carlo simulations (300 runs, each run with T = 5000 steps) with different approaches to define the value of ε.

same holds true for the enthalpy of complexation, where the average parameter is $-4.621\ kcal/mol$ and the average IPR is 0.057. The estimated parameters from *pytc* and *SupraFit* are the same.

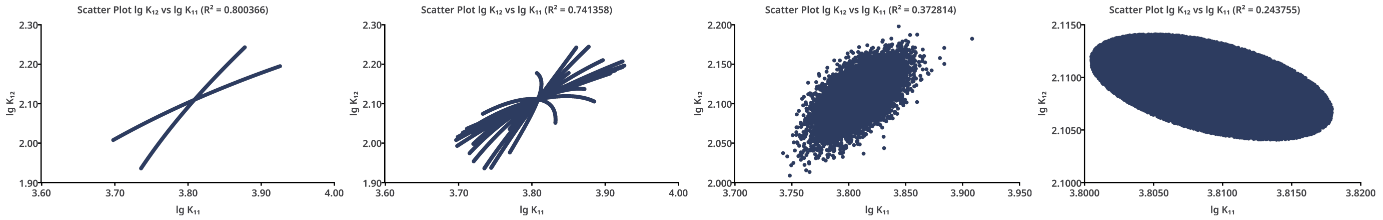
Table 12 Estimated parameters with *pytc* and *SupraFit* for *hepes-01*, *hepes-02* and *hepes-03* and the global models with the 95% confidence intervals. In *SupraFit*, MC derived confidence intervals were obtained using $\sigma_{MC} = SE_y$ and 5000 steps. The IPR is given in round brackets.

	lgK_{11}	$[lgK_{11,-}, lgK_{11,+}]$	ΔH_{AB}	$[\Delta H_{AB,-}, \Delta H_{AB,+}]$
				$\frac{kcal}{mol}$
pytc	7.594	[7.580 , 7.607]	-4.621	[-4.633 , -4.610]
MM	7.594	[7.573 , 7.614] (0.041)	-4.621	[-4.640 , -4.603] (0.037)
01	7.567	[7.546 , 7.587] (0.040)	-4.613	[-4.630 , -4.595] (0.035)
02	7.604	[7.562 , 7.646] (0.084)	-4.668	[-4.706 , -4.630] (0.076)
03	7.614	[7.579 , 7.651] (0.072)	-4.582	[-4.612 , -4.553] (0.059)

Dilution

The same example data set from *pytc* was used to analyse the effect of the dilution experiments on the parameter estimation. The four approaches, described in section 3.2.3, were applied: As first approach (1) the titration was analysed with dilution correction, included according to equation 10 but without referring to any external blank titration. Including dilution using another experiment was realised as follows: (2) An external blank titration was used to estimate the two dilution parameters m_δ and n_δ in equation 10, which were included and kept constants while lgK_{11}, ΔH and fx were obtained. The third parameter estimation (3) was performed using equation 8 after the blank experiment was subtracted from the complexation experiment. In the last experiment (4) the dilution and the complexation experiment were combined as *MetaModel*. Therefore m_δ and n_δ were estimated using the dilution and the titration experiment globally, while lgK_{11}, ΔH and fx were estimated locally, using only the data from the titration experiment. The corresponding isotherm and blank experiment are shown in Fig. 19, the estimated parameter for *hepes-01* are listed in Table 13. The heat observed from the dilution experiment is very small, compared to the heat from binding experiment. Fig. S15 contains the three isotherms and dilution experiments for the *hepes-01*, *imid-01* and *tris-01* data sets. See Tables S3 - S5 for all best fit values as well as the confidence intervals of the parameters lgK_{11} and ΔH_{AB}.

Monte Carlo simulation with 20000 steps and $\sigma_{MC} = SE_y$ were performed, the corresponding boxplots for lgK_{11} and ΔH_{AB} in case



(a) Weakened Grid Search with only lgK_{11} and lgK_{12} included, $R^2 = 0.80$

(b) Weakened Grid Search with all parameters included, $R^2 = 0.74$

(c) Monte Carlo simulation, $R^2 = 0.37$

(d) Model Comparison with only lgK_{11} and lgK_{12} included, $R^2 = 0.24$

Fig. 17 The resulting scatter plots for lgK_{11} and lgK_{12} differ for various statistical approaches.

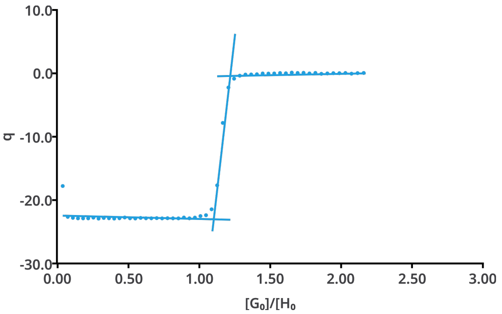
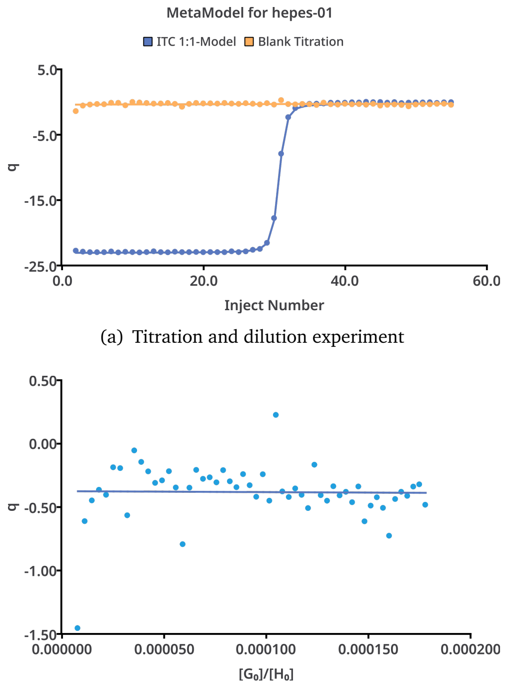


Fig. 18 The initial value for fx is guessed using three linear functions.



(a) Titration and dilution experiment

(b) Linear fit for the dilution experiment, $R^2 = 2.88 \cdot 10^{-4}$

Fig. 19 Isotherms for the complexation experiments and the dilution. Data are taken from *hepes-01* of the *pytc-demo.* [16]

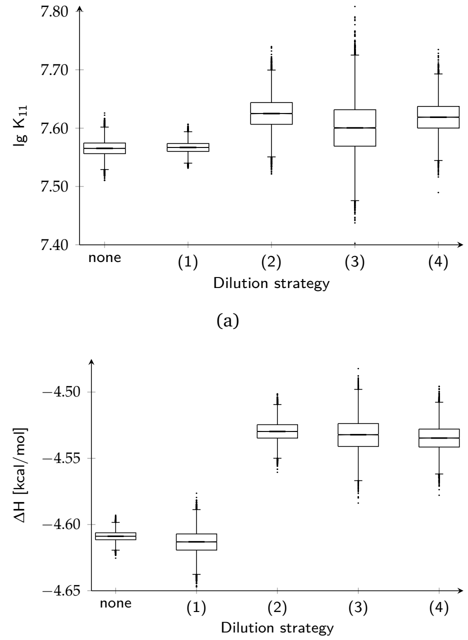
Table 13 Estimated parameters lgK_{11} and ΔH with the IPR and standard deviation of the confidence intervals calculated via Monte Carlo simulation using *hepes-01* data set and different dilution strategies

	lgK_{11}			ΔH $[kcal/mol]$		
		IPR	σ		IPR	$\sigma \cdot 10^{-3}$
none	7.565	0.054	0.014	-4.608	0.016	3.939
(1)	7.567	0.039	0.010	-4.613	0.035	9.037
(2)	7.625	0.110	0.028	-4.529	0.029	7.593
(3)	7.599	0.180	0.046	-4.532	0.051	12.822
(4)	7.619	0.108	0.027	-4.534	0.039	10.110

of *hepes-01* are shown in Fig. 20. Boxplots including all parameters and the data sets *imid-01* and *tris-01* can be found in the supplementary information in Fig. S16 - S20.

In the *hepes-01* data set, the differences between neglecting the dilution and strategy (1) are very small in case of the estimated values for lgK_{11} and ΔH_{AB}. However, Monte Carlo simulations reveal, that there is an influence on the confidence intervals. For both, *imid* and *tris* data, the differences between the estimated parameters (lgK_{11} and ΔH) and the corresponding confidence intervals comparing the neglected dilution and strategy (1) are much more intense (see Fig. S16 and S17). The results after explicitly including the dilution experiment in the parameter estimation following the three remaining approaches show that all three methods result in different best-fit parameters as compared to none dilution and strategy (1). However, the Monte Carlo simulations indicate, that the subtraction of the results of the *hepes* blank experiment deteriorates the statistical parameters of the obtained values for lgK_{11} and ΔH compared to strategy (2) and (4). In the *imid* and *tris* data sets, similar broadened confidence intervals as indicated by IPR and σ were not observed (see Table S16 and Fig. S17). This can be explained by using the correlation coefficient for the linear fit of the dilution experiment (Fig. 19(b) and Fig. S15), where R^2 is worst for *hepes* ($R^2 = 2.88 \cdot 10^{-4}$) and better for *imid* ($R^2 = 1.58 \cdot 10^{-3}$) and *tris* ($R^2 = 6.52 \cdot 10^{-3}$) dilution data.

It was demonstrated, how the influence of various approaches to include blank experiments can be analysed using Monte Carlo simulation. In the present example, the obtained parameters only change on a very small scale, e.g. the heat of complexation varies in scales of less than 0.5 kcal/mol due to the small heat of dilution. This may however not be true in general and statistical post-processing can help to understand the obtained results more deeply.



(a)

(b)

Fig. 20 Boxplot of (a) lgK_{11} and (b) ΔH values obtained from Monte Carlo simulations performed on the *hepes-01* data sets with different dilution strategies tested.

6 Conclusion

A new graphical program to perform non-linear regression with focus on the calculation of stability constants by means of NMR titration and ITC experiments has been presented. The software is written in C++, using the Qt Toolkit and the Eigen library and is fully open source and therefore transparent regarding the underlying mathematics and algorithms. Additionally to the pure estimation of the various physical parameters, that are used to describe the complexation process, statistical analysis can be performed to obtain confidence intervals for each single parameter and to gain a deeper insight in the performed experiments. The adoption of several techniques are reported, which are already described in the literature (Monte Carlo simulation and F-Test approaches), however the routinely usage of these approaches has not been reported yet. We hope, that *SupraFit* provides a good basis to analyse titration experiments with respect to the statistical judgement and to further improve the insight in the supramolecular systems. We additionally aim to provide *SupraFit* as easy-as-necessary and as powerfull-as-possible regarding the usability of the user interface, that all the tools brought by *SupraFit* are straightforwardly accessible. Contributions like new models or statistical post-processing are welcome.

The source code and binaries of *SupraFit* can be obtained free of charge from the GitHub repository at https://github.com/conradhuebler/SupraFit.

Conflicts

There are no conflicts to declare.

Acknowledgement

The author thanks Prof. M. Mazik, TU Bergakademie Freiberg for her support as well as Dr. Sebastian Förster and Dr. Stefan Kaiser for finding bugs and constructive feedback on *SupraFit* and Dr. Jürgen Seidel for helpful his feedback on that manuscript and Mara Büßemeyer for proofreading. C.H gratefully acknowledges the Centre of Advanced Study and Research - Freiberg (GraFA) for funding.

Appendix

Appendix A: abbreviation and symbols

$[A]$	concentration of component A
$[A]_0$	initial concentration of component A
$[B]$	concentration of component B
$[B]_0$	initial concentration of component B
$[X]$	concentration of any component
K_{11}	step-wise stability constant for a 1:1 complex
K_{21}	step-wise stability constant for a 2:1 complex
K_{12}	step-wise stability constant for a 1:2 complex
β_{21}	stability constant for a 2:1 system
β_{12}	stability constant for a 1:2 system
y	observed signal or physical property, dependent data
Y	proportionality factor linking concentration with y
δ	observed chemical shift
A_{abs}	observed absorbance
ε_i	extinction coefficient
V	cell volume
v	inject volume
Q	observed heat
ΔH	heat of formation
m_δ, n_δ	linear coefficients in blank experiments
E	enzyme
S	substrate
P	product
K_M	Michaelis-Menten constant
v_{max}	maximum reaction rate
r	reaction rate
θ	parameter in general
$\hat{\theta}$	estimated parameter / best-fit parameter
$\tilde{\theta}$	true value
$[\theta_-, \theta_+]$	confidence interval, range within $\tilde{\theta}$ is expected to be
IPR	inter-percentile range
x	independent data
y_{exp}	experimental data
y_{calc}	(re)calculated experimental data using $\hat{\theta}$
SSE	sum of squared errors
e	residual, error: $(y_{exp} - y_{calc})$
ε	random error
σ_{fit}	standard deviation of the residuals
SE_y	standard error
χ^2	chi-squared error
T	number of Monte Carlo steps
σ_{MC}	standard deviation used to set up Monte Carlo simulations
$N(\mu, \sigma^2)$	normal distribution with mean μ and standard deviation σ
μ	mean of normal distribution
σ	standard deviation of normal distribution
α	probability
K	number of parameters
N	number of data points
$F_{N,N-K}$	critical value in the F-distribution
δ_θ	increment to change θ during WGS and MOC
WGS	Weakened Grid Search
MOC	Model Comparison
MC	Monte Carlo simulation
BS	Bootstrapping

Appendix B: Equilibrium equations

Systems of 1:1 stoichiometry

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB} \tag{38}$$

$$K_{11} = \beta_{11} = \frac{[AB]}{[A][B]} \tag{39}$$

$$[A]_0 = [A] + \beta_{11}[A][B] = [A] + [AB] \tag{40}$$

$$[B]_0 = [B] + \beta_{11}[A][B] = [B] + [AB] \tag{41}$$

$$K_{11} = \frac{[AB]}{([A]_0 - [AB]) \cdot ([B]_0 - [AB])} \tag{42}$$

$$0 = K_{11}([A]_0 - [AB]) \cdot ([B]_0 - [AB])) - [AB] \tag{43}$$

$$0 = K_{11}[AB]^2 - [AB](K_{11}[A]_0 + K_{11}[B]_0 + 1) + K_{11}[A]_0[B]_0 \tag{44}$$

Systems of 2:1/1:1 stoichiometry

With the mass balance equations

$$\begin{aligned} [A]_0 &= [A] + [AB] + 2[A_2B] \\ &= [A] + \beta_{11}[A][B] + 2\beta_{21}[A]^2[B] \end{aligned} \tag{45}$$

$$\begin{aligned} [B]_0 &= [B] + [AB] + [A_2B] \\ &= [B] + \beta_{11}[A][B] + \beta_{21}[A]^2[B] \end{aligned} \tag{46}$$

follows the concentration of unbound host: [15]

$$0 = [A]^3A + [A]^2B + [A]C - [A]_0 \tag{47}$$

$$A = K_{11}K_{21}$$

$$B = K_{11}(2K_{21}[B]_0 - K_{21}[A]_0 + 1)$$

$$C = K_{11}([B]_0 - [A]_0) + 1$$

Systems of 1:1/1:2 stoichiometry

The mass balance equations are formed similarly to the other systems, with $\beta_{12} = K_{11}K_{12}$

$$\begin{aligned}[A]_0 &= [A] + [AB] + [AB_2] \\ &= [A] + K_{11}[A][B] + K_{11}K_{12}[A][B]^2 \\ &= [A] + \beta_{11}[A][B] + \beta_{12}[A][B]^2\end{aligned} \tag{48}$$

$$\begin{aligned}[B]_0 &= [B] + [AB] + 2[AB_2] \\ &= [B] + K_{11}[A][B] + 2K_{11}K_{12}[A][B]^2 \\ &= [B] + \beta_{11}[A][B] + 2\beta_{12}[A][B]^2\end{aligned} \tag{49}$$

$$\begin{aligned}0 &= [B]^3A + [B]^2B + [B]C - [B]_0 \\ A &= K_{11}K_{12} \\ B &= K_{11}(2K_{12}[A]_0 - K_{12}[B]_0 + 1) \\ C &= K_{11}([A]_0 - [B]_0) + 1\end{aligned} \tag{50}$$

Systems of 2:1/1:1/1:2 stoichiometry

The solution of that system is defined by the mass-balance equation

$$\begin{aligned}[A]_0 &= [A] + [AB] + [AB_2] + 2[A_2B] \\ &= [A] + K_{11}[A][B] + K_{11}K_{12}[A][B]^2 + 2K_{21}K_{11}[A]^2[B] \\ &= [A] + \beta_{11}[A][B] + \beta_{12}[A][B]^2 + 2\beta_{21}[A]^2[B]\end{aligned} \tag{51}$$

$$\begin{aligned}[B]_0 &= [B] + [AB] + 2[AB_2] + [A_2B] \\ &= [B] + K_{11}[A][B] + 2K_{11}K_{12}[A][B]^2 + K_{21}K_{11}[A]^2[B] \\ &= [B] + \beta_{11}[A][B] + 2\beta_{12}[A][B]^2 + \beta_{21}[A]^2[B]\end{aligned} \tag{52}$$

References

1 *Qt-Toolkit: https://www.qt.io/ 17.01.2022*.

2 Guennebaud, Gaël and Jacob, Benoît and others, *Eigen v3*, http://eigen.tuxfamily.org, 2010.

3 *GNU General Public License: http://www.gnu.org/licenses/gpl.html 17.01.2022*.

4 *The MIT License: https://opensource.org/licenses/mit-license.php 17.01.2022*.

5 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminf.*, 2012, **4**, 17.

6 M. Brehm and B. Kirchner, *J Chem Inf Model*, 2011, **51**, 2007–2023.

7 M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus *et al.*, *Comput. Phys. Comm.*, 2010, **181**, 14771489.

8 N. M. OBoyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminf.*, 2011, **3**, 114.

9 M. J. Hynes, *J. Chem. Soc., Dalton Trans.*, 1993, 311–312.

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

11 D. B. Hibbert and P. Thordarson, *Chem. Commun.*, 2016, **52**, 12792–12805.

12 P. Thordarson, *supramolecular.org*, http://supramolecular.org, 2019.

13 P. Thordarson, *opennanomed*, http://opennanomed.org/, 2017.

14 P. Thordarson, *OpenKinetics*, http://openkinetics.org/, 2017.

15 P. Thordarson, *Chem. Soc. Rev.*, 2011, **40**, 13051323.

16 H. Duvvuri, L. C. Wheeler and M. J. Harms, *Biochemistry*, 2018, **57**, 2578–2583.

17 H. Akaike, *Trans. Autom. Control*, 1974, **19**, 716–723.

18 N. Sugiura, *Commun. Stat. - Theory Methods*, 1978, **7**, 13–26.

19 C. M. HURVICH and C.-L. TSAI, *Biometrika*, 1989, **76**, 297–307.

20 G. Schwarz, *Ann. Stat.*, 1978, **6**, 461–464.

21 H. Zhao, G. Piszczek and P. Schuck, *Methods*, 2015, **76**, 137 – 148.

22 S. Keller, C. Vargas, H. Zhao, G. Piszczek, C. A. Brautigam and P. Schuck, *Anal. Chem*, 2012, **84**, 5066–5073.

23 P. Thordarson, *Techniques in Supramolecular Chemistry : From Molecules to Nanomaterials*, WILEY-VCH Verlag GmbH & Co. KGaA, 2012, vol. 2.

24 B. Valeur, M. N. Berberan-Santos and M. M. Martin, in *Photophysics and Photochemistry of Supramolecular Systems*, John Wiley & Sons, Ltd, 2006, ch. 7, pp. 220–264.

25 K. Hirose, in *Quantitative Analysis of Binding Properties*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2012, ch. 2, p. 2766.

26 E. Freire, O. L. Mayorga and M. Straume, *Anal. Chem*, 1990, **62**, 950A–959A.

27 E. Freire, A. Schön and A. Velazquez-Campoy, *Meth. Enzymol.*, 2009, **455**, 127–155.

28 F. P. Schmidtchen, in *Isothermal Titration Calorimetry in Supramolecular Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2012, ch. 3, p. 67103.

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

30 O. Francesconi, M. Martinucci, L. Badii, C. Nativi and S. Roelens, *Chem. Eur. J.*, 2018, **24**, 6828–6836.

31 L. Alderighi, P. Gans, A. Ienco, D. Peters, A. Sabatini and A. Vacca, *Coord. Chem. Rev.*, 1999, **184**, 311–318.

32 A. S. Mahadevi and G. N. Sastry, *Chem. Rev.*, 2016, **116**, 2775–2825.

33 L. Tebben, C. Mück-Lichtenfeld, G. Fernández, S. Grimme and A. Studer, *Chem. Eur. J.*, 2017, **23**, 5864 – 5873.

34 J. F. Rusling and T. F. Kumosinski, *Nonlinear computer modeling of chemical and biochemical data*, Academic Press, 1996.

35 K. Levenberg, *Quarterly of applied mathematics*, 1944, **2**, 164168.

36 D. W. Marquardt, *SIAM Journal on Applied Mathematics*, 1963, **11**, 431441.

37 C. Hübler, *conradhuebler/SupraFit*, 2019, `https://doi.org/10.5281/zenodo.3364569`.

38 B. Efron, *The jackknife, the bootstrap, and other resampling plans*, Siam, 1982, vol. 38.

39 P. Bevington, *Data Reduction and Error Analysis for Physicists*, 1969.

40 R. E. Barrans Jr and D. A. Dougherty, *Supramol Chem*, 1994, **4**, 121–130.

41 J. Tellinghuisen, *Anal. Biochem.*, 2003, **321**, 79–88.

42 J. Tellinghuisen, *J. Phys. Chem. B*, 2005, **109**, 20027–20035.

43 H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data using Linear and Nonlinear Regression. A practical guide to curve fitting.*, GraphPad Software Inc., www.graphpad.com, 2003.

44 B. Efron, *Ann. Statist.*, 1979, **7**, 1–26.

45 A. J. Canty, A. C. Davison, D. V. Hinkley and V. Ventura, *Can J Stat*, 2006, **34**, 5–27.

46 B. Efron and T. Hastie, *Computer age statistical inference*, Cambridge University Press, 2016, vol. 5.

47 G. E. P. Box, *Ann. N. Y. Acad. Sci.*, 1960, **86**, 792–816.

48 E. M. L. Beale, *J. R. Stat. Soc. Series. B Stat. Methodol.*, 1960, **22**, 41–76.

49 J. M. Beechem, *Numerical Computer Methods*, Academic Press, 1992, vol. 210, pp. 37 – 54.

50 D. Bates and D. Watts, *Nonlinear Regression Analysis and Its Applications*, Wiley, 1988.

51 K. Vugrin, L. Swiler, R. Roberts, N. Stucky-Mack and S. Sullivan, *Water Resour. Res.*, 2007, **43**, year.

52 G. Kemmer and S. Keller, *Nat. Protoc.*, 2010, **5**, 267281.

53 P. Gramatica, *QSAR Comb Sci*, 2007, **26**, 694–701.

54 The Gnome Project, *The Gnumeric Spreadsheet: Free, Fast, Accurate — pick any three v.1.12.41*, http://www.gnumeric.org/, 2018.

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