Determining association constants from titration experiments in supramolecular chemistry†

Pall Thordarson*

Received 10th August 2010
DOI: 10.1039/c0cs00062k

The most common approach for quantifying interactions in supramolecular chemistry is a titration of the guest to solution of the host, noting the changes in some physical property through NMR, UV-Vis, fluorescence or other techniques. Despite the apparent simplicity of this approach, there are several issues that need to be carefully addressed to ensure that the final results are reliable. This includes the use of non-linear rather than linear regression methods, careful choice of stoichiometric binding model, the choice of method (*e.g.*, NMR *vs.* UV-Vis) and concentration of host, the application of advanced data analysis methods such as global analysis and finally the estimation of uncertainties and confidence intervals for the results obtained. This *tutorial review* will give a systematic overview of all these issues—highlighting some of the key messages herein with simulated data analysis examples.

1. Introduction

The field of modern supramolecular chemistry continues to grow from strength to strength from its humble beginnings in 1960–1970 with the pioneering work of Pedersen, Lehn and Cram that was subsequently recognised by the Nobel prize in 1987.[1–3] One of the fundamental issues in supramolecular chemistry has always been the quantitative analysis of the intermolecular interactions of interest. The most common approach to address this issue is the supramolecular titration method. Here, one component (guest) is gradually added to the system (host) while monitoring a physical property such as specific chemical resonance (NMR) or absorption band (UV) that is sensitive to the supramolecular interaction(s) of interest. The resulting information is then compared and fitted to binding models to obtain information such as the association constant K_a (See Appendix A for abbreviations used in this *tutorial review*.), energetics (ΔG, ΔH and ΔS) and stoichiometry (1 : 1 *vs.* 1 : 2 *etc*).

The very first crown-ether reported by Pedersen was the dicyclohexyl-18-crown-6 (Fig. 1a, n.b., at the time of this work the exact stereochemistry of this host was not defined).[4] Although a lot of the subsequent work was done with ion-selective electrodes, the first report on the association constant for *cis*-dicyclohexyl-18-crown-6 binding of potassium (K^+) and other alkali metals applied calorimetric titration to obtain K_a = 100 for K^+ (solvent, temperature and counterion were not specified).[5] A recent, much more complex example for

School of Chemistry, The University of New South Wales, Sydney, NSW 2052, Australia. E-mail: p.thordarson@unsw.edu.au; Fax: +61 29385 4478; Tel: +61 29385 4478

† Electronic supplementary information (ESI) available: Matlab *m*-files for analysing and simulating 1 : 1, 1 : 2 and 2 : 1 equilibria for NMR and UV-Vis titration, including example data files. See DOI: 10.1039/c0cs00062k

Pall Thordarson

Pall Thordarson obtained his BSc from the University of Iceland in 1996 and a PhD from The University of Sydney in 2001, followed by a Marie Curie Fellowship at the University of Nijmegen, The Netherlands. He returned to Australia in 2003 and obtained an ARC Australian Research Fellow at The University of Sydney in 2006. He was appointed as a Senior Lecturer at the University of New South Wales (UNSW), Sydney, Australia, in 2007 where he leads a research group working on light-activated bioconjugates for controlling enzyme activity, self-assembled gels for drug delivery and supramolecular chemistry.

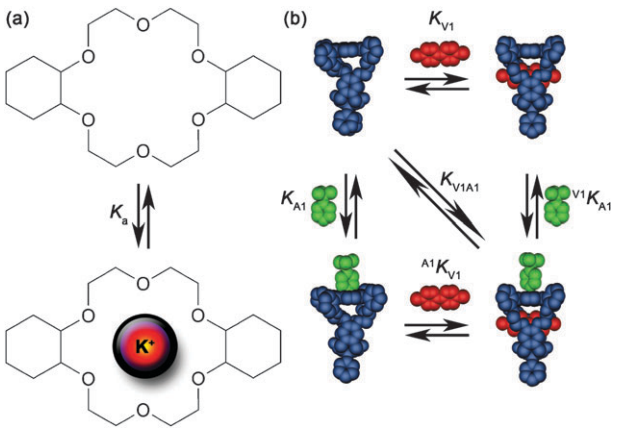

Fig. 1 Typical supramolecular equilibria. (a) Pedersen's dicyclohexyl-18-crown-6 (the stereochemistry was not specified) binding to a potassium ion.[4,5] (b) Squaring cooperative cycles with porphyrin clip host (blue), viologen (red) and *tert*-butyl pyridine (green).[6]

what sort of data can be extracted from relatively simple titration studies (fluorescence) comes from Rowan and co-workers in their study on "squaring cooperative cycles" (Fig. 1b).[6] The bulk of the current literature employing supramolecular titration is somewhere between the two above extremes in terms of simplicity *vs.* complexity, with the emphasis shifting towards the latter.

The quality of the information gained from supramolecular titrations depends on a number of factors. These include of course operational factors such as purity of compounds used or accuracy of the measurement devices used. However, the planning, execution and subsequent data analysis are as important. Careful consideration of all these factors is the key to obtaining quality information that can then be used to guide further research into the application of the systems. The above mentioned example of squaring cooperative cycles is a good example—without proper quantitative data, the application of cooperative interactions to direct self-assembly[7] becomes an empirical rather than predictable science. Many excellent textbooks and reviews deal with some of these issues, such as the much quoted monograph by Connors on binding constants.[8] That said, many researchers still get caught in avoidable traps such as using out-dated linear regression methods in data analysis or over-reliance on the Job's method to determine stoichiometry, to name just two examples of other important issues in the determination of association constants from supramolecular titration. This *tutorial review* aims to help people avoid these traps and identify the key issues involved in planning, executing and analysing data from supramolecular titration experiments. To simplify the discussion the emphasis will be on relatively simple host–guest equilibria such as the formation of 1 : 1, 1 : 2 and 2 : 1 complexes but the lessons learned from these systems can also be applied to more complex systems such as the cooperative aggregation of monomers into oligomers.[9] The review will start by outlining some of the key dos and don'ts of supramolecular titrations before moving on to a more detailed description of equations and equilibria, determination of stoichiometry, common techniques and data analysis.

2. Dos and don'ts of supramolecular titration experiments

At the outset of this *tutorial review* it is useful to list some of key messages contained herein in the form of fourteen dos and don'ts:

1. Do try to obtain as much information as possible *prior* to carrying out the titration experiment on what the expected outcomes are. Look at related or similar systems for clues. Starting with a reasonable estimation of what the stoichiometry and the association are likely to be will make it easier to choose the right technique, concentration of host/guest, number of addition points, *etc*. If little or no helpful information exists on the system of interest a quick preliminary titration experiment might be worthwhile.

2. Don't just simply apply a "black box" binding model from a computer program or a book without trying to understand how they are derived from the underlying equilibria. Understanding the link between binding equilibria and the equations used to obtain binding constants helps with evaluating how robust the results are and whether the right model has been used to analyse the data.

3. Do take the extra time necessary to carry out the actual titration experiment with as much care as possible. Check the purity of chemicals used and thoroughly clean any glassware used. Use the most accurate methods available to measure out the solution—often weighing out the solvents for solutions is much more accurate than relying simply on volumetric glassware. Avoid dilution factors, *e.g.*, by preparing the guest solutions in the host solution with the same concentration of the host for the titration. This will ensure the host concentration stays constant during the course of the experiment cancelling (mostly) out any additional effects from aggregation of the host. Make sure to control and record the temperature. An accurate association constant at an unknown temperature is of little value.

4. Don't choose an instrumental method to follow the course of titration on economical (I don't need deuterated solvents for UV-Vis) or emotional (I really like doing NMR) grounds but try to pick the method that is most applicable to the system of interest. Admittedly, economical reasons do sometimes play a role (*e.g.*, access to calorimeters is not universal) but the primary concern should be factors such as concentration *vs.* expected association strength (NMR works well with $K_a < 100\,000\ M^{-1}$), nature of host/guest used (*e.g.*, chromophores) or the potential influences of impurities (fluorescence spectroscopy can be more sensitive to impurities than UV-Vis spectroscopy).

5. Do aim to get a reasonable number of titration points (additions) as practically feasible in a reasonable short time. The extra time spent getting a reasonable number will ensure better certainty in the data fitting process, especially when strong interactions or more complex equilibria such as 1 : 2 host/guest complexation can play a role. Note that collecting a large number points can be counterproductive due to building of systematic errors over time (*e.g.*, evaporation of solvent). On balance it is better to collect 3 titrations with 15 data points than one with 45 points.

6. Don't take any shortcuts in the data analysis, especially by using out-dated linear regression methods such as the Benesi–Hildebrand/Lineweaver–Burk, Scott/Hanse–Woolf or Scatchard transformations. These methods were useful before modern computers and programs became available but they introduce a number of errors and problems that can be avoided with modern computer-based non-linear regression methods. Related to this is the problem of making assumptions such as equating the total concentration of the guest ($[G]_0$) with the free (unbound) concentration of the guest [G] or that the measured quantity (*e.g.*, absorbance change ΔA) equals the change in the concentration of the complex ([HG]). Frequently, these problems can be traced back to using approaches that still work well in studying association phenomena in biochemistry but are actually not applicable to typical supramolecular chemistry titrations. Finally, when analysing 1 : 2 or more complex equilibria, use programs that can handle the underlying cubic or more complex equations rather than making unnecessary and risky shortcuts.

7. Do consider all sensible possible stoichiometries and use as many techniques as possible to verify or otherwise a possible stoichiometric model.

8. Don't rely exclusively on one particular method to determine stoichiometry such as the Job's method which really only works well when there is one type of complex present—usually the 1 : 1. If possible and at least remotely probable, the data should be analysed for more than one type of stoichiometry and the results then evaluated (see Section 4 below).

9. Do consider the method of global analysis or simultaneous fitting of multiple datasets to a single binding model. Generally this improves the fitting processes by tightening the error surfaces and thus increasing the likelihood that the solution found in the regression process is the true and not a local minimum.

10. Don't forget to critically compare different data analysis (fitting) results in terms of the number of parameters, *etc*. used. With a high enough number of parameters, any data can be fitted accurately. Look also beyond the association constant obtained; if other parameters obtained are nonsensical, such as a chemical shift difference of +120 ppm in ^{1}H NMR studies, the model used is probably invalid. Also, if the fitting results are "unstable" in terms of sensitivity to the initial guess of parameters, the results are probably not reliable.

11. Do repeat the titration. A value from a single run titration should not be published. Repeating titration is invaluable when it comes to estimating uncertainties.

12. Don't just accept the output of the data analysis program; think about the shape and information content of the binding isotherm and the residual plot.

13. Do try to estimate the uncertainty in your final results. Estimating confidence intervals based on asymptoting standard errors does sometimes work but this method needs to be used with great care as it is based on assumption that strictly don't apply to non-linear regression methods. Additionally, this popular (and other related classical) statistical methods ignore uncertainties in factors such as the host concentration which often is the largest one in the system. Other approaches can be used (*e.g*., Monte Carlo) but in the end nothing really replaces repeating the experiment several times to obtain some sort of a statistically meaningful estimation of uncertainty in the results obtained.

14. Don't focus just on the association constants obtained when interpreting the results and comparing with other systems. Think also about the results in terms of energy: $\Delta G = -RT\ln K_a$. Doubling an association constant increases ΔG by -1.7 kJ mol^{-1}—this equates to 18% increase in energy in going from K_a of 50 M^{-1} to 100 M^{-1} whereas for K_a = 500 000 M^{-1}, doubling the association constants only increases the ΔG by 5%. Additionally, one needs to consider statistical factors when analysing results from multivalent interactions, *e.g*., in the case of 1 : 2 complexation where $K_1 = 4K_2$ if there is no cooperativity in the system.

The above list is not exhaustive but it should help with ensuring that the final outcomes of a supramolecular titration experiment are reliable for interpretation. The more detailed discussion below will expand on these points—comparing approaches and techniques and highlighting some of the issues with simulations of typical titration data.

3. Equations and equilibria—quantifying the extent of complexation

The below discussion is largely based on previous textbooks and reviews on determining binding constants.[8,10,11] To facilitate better understanding of the remainder of this *tutorial review* for newcomers to the field, the derivation of the key equations is summarised here again, focusing on 1 : 1 equilibria. A summary of all the key equations for 1 : 1, 1 : 2 and 2 : 1 equilibria is provided in Charts 1–3—this might also be useful for readers who are quite familiar with these equations.

3.1 The basics—equilibria and association constants

With the aid of computer programs designed for determining association constants from titration experiments, one might be tempted to ignore the underlying equilibria and the associated mathematical equations, and just report the answers obtained from these programs. This "black box" approach is, however, unwise as without this understanding it becomes very difficult to answer such questions as: (i) did I determine the stoichiometry correctly, (ii) did I use the right binding model and how I can compare different models, (iii) did I use the right technique and concentration(s) of the host/guest used and finally (iv) how reliable are the results obtained? The first step toward answering these questions is to look at how the typical equilibria in supramolecular chemistry is defined, starting with the simple 1 : 1 equilibria (Fig. 2a) according to eqn (1).[8,10,11]

$$K_a = \frac{[HG]}{[H][G]} \quad (1)$$

The choice of "host" and "guest" is somewhat arbitrary but in this *tutorial review* the host concentration is the one that is kept constant while being titrated with the guest. For practical reasons, the host is usually the larger, more expensive (or more synthetically challenging) component in these systems. It should be pointed out that other textbooks and references may use the different terms such as ligand, receptor or the substrate for either (or both) the host or the guest as they have been defined here. The thermodynamic association constant K_x is related to kinetics of the particular system, hence $K_a = k_1/k_{-1}$, with the k_1 and k_{-1} the forward (on) and backward (off) rate constants for the equilibria of interest. If these rate constants can be measured, the association constant K_a can be calculated directly, but these kinetic methods are outside the scope of this review.

The 1 : 1 equilibria is just a special case of general equation for the formation of any supramolecular m × host·n × guest complex (Fig. 2b) according to eqn (2).[10]

$$\beta_{mn} = \frac{[H_mG_n]}{[H]^m[G]^n} \quad (2)$$

When $m = n = 1$ we arrive back at eqn (1) while $m = 1$ and $n = 2$ describes the fairly common 1 : 2 equilibria (Fig. 2c) for two guests bound to a host. It would be extremely unlikely for such a termolecular complex to form by simultaneous collision of one host with two guest—a step wise process as described in

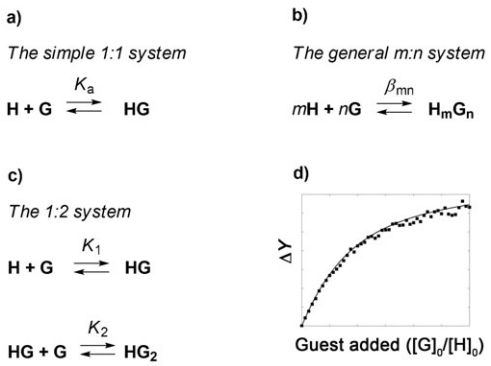

Fig. 2 (a)–(c) Three of the supramolecular equilibria discussed in this review: **H** = host, **G** = Guest and $\mathbf{H}_x\mathbf{G}_x$ = host–guest complex of interest, K_x = the thermodynamic association constant for a particular interest and β_{mn} = overall association constant for an $m : n$ host–guest complex formation. (d) A typical binding isotherm. See text for details.

eqn (3) and (4) is a more plausible description of how such a complex is formed.[10,11]

$$K_1 = \frac{[HG]}{[H][G]} \quad (3)$$

$$K_2 = \frac{[HG_2]}{[H][HG]} \quad (4)$$

The relationship between these stepwise constants and their physical meaning will be described in further details below after methods to determine association constants for 1 : 1 systems have been described.

3.2 The archetypical 1 : 1 system

Exactly how do we quantify [H], [G] and [HG] in eqn (1)? In biochemistry it is sometimes possible to measure directly the concentration of the complex [HG] (or [H] or [G]) by methods such as electrophoresis.[12] If the concentration of the complex [HG] is known the remaining concentrations can be determined by the mass balance as described by eqn (5) and (6) for simple 1 : 1 binding.

$$[H]_0 = [H] + [HG] \quad (5)$$

$$[G]_0 = [G] + [HG] \quad (6)$$

Unfortunately, it is usually not possible to measure [HG] (or [H] and [G]) in supramolecular chemistry experiments, but the knowledge of these is required to determine K_a according to eqn (1). The only notable exception is when the free host and guest are in true slow exchange with their complexes on the NMR timescale so that two distinct peaks can be observed. In this case it is in principle possible to obtain the concentration(s) of the species involved by integration but as discussed below, this approach is confounded by a number of issues, limiting its appeal even further. Thankfully, there are other accurate methods available for the determination of K_a—methods that rely on indirectly determining the concentration of the complex [HG] from titration experiments. Supramolecular titration experiments are usually carried out by fixing the concentration of one component, the host (**H**), while the concentration of another component, the guest (**G**), is varied (this has also been called the dilution method). During the course of this titration, the physical changes in the system are monitored, usually spectroscopically, and this change is then plotted as a function of guest added to host (equivalent guests). The resulting titration curve known as *a binding isotherm* is then fitted to a mathematical model that is derived from the *assumed equilibria* to obtain the association constant (K_a) and/or other physical constants of interest.

The mathematical model used to obtain the association constant is usually developed from realising that the physical change (ΔY, *e.g.*, a NMR shift or an change in UV-Vis absorbance) observed is correlated to the concentration of the complex [HG] as $\Delta Y \propto [HG]$, or in some cases, the free host [H] or the free guest [G]. The physical change (Y) being monitored can usually be described as the aggregate of the individual components according to eqn (7) as a function of concentration (*e.g.*, for UV-Vis spectroscopy) or eqn (8) as a function of mole fractions f_X (f_X defined as: $f_X = [X]/[X]_0$) in the special case of NMR.

$$Y = Y_H[H] + Y_G[G] + Y_{HG}[HG] \quad (7)$$

$$Y = Y_H f_H + Y_G f_G + Y_{HG} f_{HG} \quad (8)$$

The mole fraction notion is particularly helpful as when we combine it with the mass balance eqn (5) and (6) and the definition of the association constant in eqn (1), we obtain eqn (9).[8,10]

$$f_{HG} = \frac{K_a[G]}{1 + K_a[G]} \quad (9)$$

This equation describes the *general binding isotherm* for a simple 1 : 1 binding system as a hyperbolic (Fig. 2d) relationship between f_{HG} and the *free guest* ([G]) concentration. Therefore, if the free guest concentration was known, it should be quite straightforward to determine the association constant assuming the observed physical change (ΔX) is related to the mole fraction of the complex, as it usually is. Sadly, the free guest concentration [G] cannot usually be measured directly so two alternative approaches are required to solve this problem, both that result in quadratic equations containing only the *total concentrations* ($[H]_0$ and $[G]_0$) and the association constant (K_a) as the unknown:

In the first approach, we start by realising that since $f_{HG} = [HG]/[H]_0$, eqn (9) can also be written as $[HG] = [H]_0K_a[G]/(1 + K_a[G])$ which after insertion to eqn (6) yields $[G]_0 = [G] + H_0K_a[G]/(1 + K_a[G])$. Rearranging yields the quadratic eqn (10) which has only one relevant real solution according to eqn (11).[8,10]

$$G^2 - [G]\left(G_0 - H_0 - \frac{1}{K_a}\right) - \frac{G_0}{K_a} = 0 \quad (10)$$

$$[G] = \frac{1}{2}\left(G_0 - H_0 - \frac{1}{K_a}\right) - \sqrt{\left(G_0 - H_0 - \frac{1}{K_a}\right)^2 + 4\frac{G_0}{K_a}} \quad (11)$$

Alternatively, we can rearrange eqn (5) and (6) to isolate for [H] and [G], respectively, and insert these into eqn (1) to

expand it in the form of $K_a = [HG]/([H]_0 - [HL])([G]_0 - [G])$. Expanding the right-hand denominator to give $K_a = [HG]/([H]_0[G]_0 - [HL])([H]_0 + [G]_0 + [HG]^2)$ and then rearranging yields the quadratic eqn (12) and the corresponding solution in eqn (13).[11]

$$[HG]^2 - [HG]\left(G_0 + H_0 + \frac{1}{K_a}\right) - [H_0][G_0] = 0 \quad (12)$$

$$[HG] = \frac{1}{2}\left(G_0 + H_0 + \frac{1}{K_a}\right) - \sqrt{\left(G_0 + H_0 + \frac{1}{K_a}\right)^2 + 4[H_0][G_0]} \quad (13)$$

The power of eqn (13) should not be understated as we can now start to develop solutions to eqn (7) and (8) that require only the knowledge of the total (or initial) concentrations of the host and guest ($[H]_0$ and $[G]_0$) in addition to the association constant (K_a) and the physical properties (Y) that are changing (ΔY) during the course of the titration.

If we further assume that one of the components is "silent" *e.g.*, a non-absorbing free guest [G], we can simplify eqn (8) by the substitution $[H] = [H]_0 - [HG]$ from eqn (5), and as before $f_{HG} = [HG]/[H]_0$ to obtain $Y = Y_H + ([HG]/[H]_0)(Y_{HG} - Y_H)$ which can be simplified further to give eqn (14).

$$\Delta Y = Y_{\Delta HG}\left(\frac{[HG]}{[H]_0}\right) \quad (14)$$

It should be noted that even if [G] is not "silent", the complexity of the analysis is not increased significantly.

Similarly, if the physical property is related to absolute concentrations and noting that the value of the physical property in the beginning of the experiment in the absence of host is $Y_0 = Y_H[H]_0$, we can rearrange eqn (7) to give eqn (15).

$$\Delta Y = Y_{\Delta HG}([HG]) \quad (15)$$

Using [HG] from eqn (13) it is now possible to describe the expected changes from a supramolecular titration experiment from two known ($[H]_0$ and $[G]_0$) and two unknown (K_a and $Y_{\Delta HG}$) parameters in eqn (14) and (15). In practice these two unknown parameters, including the association constant of interest, are obtained by non-linear regression of the data obtained. As described in details below, this is readily achieved by various computer programs that use an algorithm and some initial guesses for K_a and $Y_{\Delta HG}$ to calculate ΔY and compare it to the measured ΔY and then vary K_a and $Y_{\Delta HG}$ until a good fit is obtained between eqn (15) and the real experimental data. The key equations for 1 : 1 equilibria are summarised in Chart 1. This includes a number of technique specific equations (NMR, UV-Vis *etc.*) that are derived directly from eqn (14) and (15) as detailed in the literature.[8,10,11]

3.3 Linear regression methods and related shortcuts

Older references and textbooks are full of examples on how some of the above expressions and equations can be simplified or transformed to linear equations ($y = a + bx$) which could then be plotted by hand to obtain the K_a and other parameters of interest by inspection of the slope and intercepts. These include the popular Benesi–Hildebrand plot ($1/f_{HG} = 1 + 1/K_a[G]$ or $1/\Delta Y = 1/Y_{\Delta HG}K_a[G] + 1/Y_{\Delta HG}$,[13] also known as the Lineweaver–Burk plot in enzyme kinetics),[14] the related Scott[15] or Hanse–Woolf[16] transformations $[G]/f_{HG} = [G] + 1/K_a$ or $[G]/\Delta Y = 1/Y_{\Delta HG}K_a + [G]/Y_0$ and the Scatchard plot ($f_{HG}/[G] = K_a - K_a f_{HG}$ or $\Delta Y/[G] = -K_a[G] + K_a Y_{\Delta HG}$)[17] but with the power of modern computers and the associated software, there is no apparent need for these transformations.

1:1 Equilibria

*General expression for equilibrium constant – **eqn (1)**:*

$$K_a = \frac{[HG]}{[H][G]}$$

*Expression for free guest concentration – **eqn (11)**:*

$$[G] = \frac{1}{2}\left(G_0 - H_0 - \frac{1}{K_a}\right) - \sqrt{\left(G_0 - H_0 - \frac{1}{K_a}\right)^2 + 4\frac{G_0}{K_a}}$$

*Expression for complex guest concentration – **eqn (13)**:*

$$[HG] = \frac{1}{2}\left(G_0 + H_0 + \frac{1}{K_a}\right) - \sqrt{\left(G_0 + H_0 + \frac{1}{K_a}\right)^2 + 4[H_0][G_0]}$$

*Change in mole fraction dependent physical property Y up on titration – **eqn (14)**:*

$$\Delta Y = Y_{\Delta HG}\left(\frac{[HG]}{[H]_0}\right)$$

*Change in absolute concentration dependent physical property Y up on titration – **eqn (15)**:*

$$\Delta Y = Y_{\Delta HG}([HG])$$

*NMR: Changes up on titration – **eqn (30)**:*

$$\Delta\delta = \delta_{\Delta HG}\left(\frac{[HG]}{[H]_0}\right)$$

*UV-Vis: Changes up on titration – **eqn (34)**:*

$$\Delta A_{obs} = \varepsilon_{\Delta HG}([HG])$$

*Fluorescence: Fluorescence of host and complex up on titration (dynamic quenching possible) – **eqn (40)**:*

$$F = k_H[H] + k_{HG}[HG] \quad (40)$$

*Fluorescence: Changes in fluorescence of host – complex up on titration (no dynamic quenching) – **eqn (42)**:*

$$\Delta F_{obs} = k_{\Delta HG}([HG])$$

*Calorimetry: Heat of formation upon on titration – **eqn (46)**:*

$$Q = \Delta H_{HG}V([HG])$$

Chart 1 Summary of equations related to 1 : 1 equilibria.

More importantly, there are two key problems associated with using these linear transformations that make their use highly questionable: (i) they violate some of the fundamental assumption of linear regression by distorting the experimental error[18,19] and (ii) they frequently involve assumptions and shortcuts such as assuming that [G] ≈ $[G]_0$ (*i.e.*, the guest is in large excess) or $Y_{HG} = Y$ at the end of titration (*i.e.*, the complex is fully formed at the end of titration which would then help to give $Y_{\Delta HG}$)—assumptions that more often than not can break down and distort the results. The *non-linear regression* approach using eqn (14) and (15) with exact solutions of the quadratic eqn (10) and (12), *produces the most accurate results*. This approach is not difficult with modern computer technology and *there is no real excuse for using old-fashion linear transformations anymore!*

The only legitimate use of linear transformations such as the Scatchard plot is to use them as an aid to visualise the results after the K_a and [G] have been calculated from non-linear regression as the human eye finds it easier to detect deviations in straight lines than variations in hyperboles.[19]

3.4 The 1 : 2 system

Having defined the 1 : 1 host–guest system, it is relatively straightforward to proceed with the more complex scenarios. The derivation of the key equations for 1 : 2 equilibria summarised in Chart 2 has been detailed previously[11] and will not be repeated here. The key equation here to note is the cubic eqn (16) as well as eqn (17) and (18) which are analogous to eqn (14) and (15) for a 1 : 1 system. The latter two equations describe the expected changes in physical properties upon titration of a host with a guest in a 1 : 2 system (assuming again that the guest is "silent", *i.e.*, $Y_G = 0$).[11] The experimental data can now be fitted by non-linear regression to eqn (17) or (18), and the underlying eqn (16), to obtain the unknown parameters K_1, K_2, $Y_{\Delta HG}$ and $Y_{\Delta HG_2}$.

The cubic eqn (16) is of special note here as it has three solutions which may or may not include complex numbers. The smallest positive real solution is the only one of relevance here and as described below, it can be obtained by certain algorithms in a number of software packages. Once the concentration of [G] is known, the concentration of [H] can also be calculated if necessary from eqn (19).[11]

$$[H] = \frac{[H]_0}{1 + K_1[G] + K_1K_2[G]^2} \qquad (19)$$

3.5 The 2 : 1 system

In principle, the host and guest can be defined at will, but if we stick to the definition that the concentration of the host is kept constant then a 2 : 1 host–guest complex would be different from the 1 : 2 complex discussed above. It is also important to distinguish the 2 : 1 system from a 1 : 2 system for the same host, *e.g.*, in the case of cyclodextrins which have been reported to form 2 : 1, 1 : 1 and 1 : 2 systems to the same guest depending on experimental conditions (making the data analysis extremely difficult).[10]

The 2 : 1 host–guest system can be described by switching **H** and **G** in eqn (3), eqn (4). The key equations, eqn (20)–(22) obtained, are summarised in Chart 3. These equations are derived in a manner similar to the 1 : 1 and 1 : 2 equilibria, noting also the relevant mass balance eqn (23) and (24).

1:2 Equilibria

*General expression for equilibrium constant – **eqn (3)** and **(4)**:*

$$K_1 = \frac{[HG]}{[H][G]} \quad (3) \qquad \text{and} \qquad K_2 = \frac{[HG_2]}{[H][HG]} \quad (4)$$

*Expression for free guest concentration – **eqn (16)**:*

$$[G]^3(A) + [G]^2(B) + [G]((C)) - [G]_0 = 0$$

With: $A = (K_1K_2)$
$B = \{K_1(2K_2[H]_0\text{-}K_2[G]_0\text{+}1)\}$
$C = \{K_1([H]_0\text{-}[G]_0)\text{+}1\}$

*Change in mole fraction dependent physical property Y up on titration – **eqn (17)**:*

$$\Delta Y = \frac{Y_{\Delta HG}K_1[G] + Y_{\Delta HG_2}K_1K_2[G]^2}{1 + K_1[G] + K_1K_2[G]^2} \quad (17)$$

*Change in absolute concentration dependent physical property Y up on titration –**eqn (18)**:*

$$\Delta Y = \frac{Y_{\Delta HG}[H]_0K_1[G] + Y_{\Delta HG_2}[H]_0K_1K_2[G]^2}{1 + K_1[G] + K_1K_2[G]^2} \quad (18)$$

*NMR: Changes up on titration – **eqn (31)**:*

$$\Delta\delta = \frac{\delta_{\Delta HG}K_1[G] + \delta_{\Delta HG_2}K_1K_2[G]^2}{1 + K_1[G] + K_1K_2[G]^2}$$

*UV-Vis: Changes up on titration – **eqn (35)**:*

$$\Delta A_{obs} = \frac{\varepsilon_{\Delta HG}[H]_0K_1[G] + \varepsilon_{\Delta HG_2}[H]_0K_1K_2[G]^2}{1 + K_1[G] + K_1K_2[G]^2}$$

Fluorescence*: Changes in fluorescence up on titration (no dynamic quenching) – **eqn (44)**:*

$$\Delta F_{obs} = \frac{k_{\Delta HG}[H]_0K_1[G] + k_{\Delta HG_2}[H]_0K_1K_2[G]^2}{1 + K_1[G] + K_1K_2[G]^2}$$

*Calorimetry: Heat of formation upon on titration – **eqn (47)**:*

$$Q = \frac{\Delta H_{HG}V[H]_0K_1[G] + \Delta H_{HG_2}V[H]_0K_1K_2[G]^2}{1 + K_1[G] + K_1K_2[G]^2}$$

Chart 2 Summary of equations related to 1 : 2 equilibria.

$$[H]_0 = [H] + [HG] + 2[H_2G] \qquad (23)$$

$$[G]_0 = [G] + [HG] + [H_2G] \qquad (24)$$

The cubic eqn (20) is analogous to eqn (16) and describes the concentration of free host [H]. The equations analogous to

2:1 Equilibria

*General expression for equilibrium constant – **eqn (3)** and **(4)**:*

$$K_1 = \frac{[HG]}{[H][G]} \quad (3) \qquad \text{and} \qquad K_2 = \frac{[H_2G]}{[G][HG]} \quad (4)$$

*Expression for free host concentration – **eqn (20)**:*

$$[H]^3(A) + [H]^2(B) + [H]((C)) - [H]_0 = 0$$

With: $A = (K_1K_2)$
$B = \{K_1(2K_2[G]_0 - K_2[H]_0 + 1)\}$
$C = \{K_1([G]_0 - [H]_0) + 1\}$

*Change in mole fraction dependent physical property Y up on titration – **eqn (21)**:*

$$\Delta Y = \frac{Y_{\Delta HG}[G]_0K_1[H] + 2Y_{\Delta H_2G}[G]_0K_1K_2[H]^2}{[H]_0\left(1 + K_1[H] + K_1K_2[H]^2\right)}$$

*Change in absolute concentration dependent physical property Y up on titration – **eqn (22)**:*

$$\Delta Y = \frac{Y_{\Delta HG}[G]_0K_1[H] + 2Y_{\Delta H_2G}[G]_0K_1K_2[H]^2}{1 + K_1[H] + K_1K_2[H]^2}$$

*NMR: Changes up on titration – **eqn (32)**:*

$$\Delta\delta = \frac{\delta_{\Delta HG}[G]_0K_1[H] + 2\delta_{\Delta H_2G}[G]_0K_1K_2[H]^2}{[H]_0\left(1 + K_1[H] + K_1K_2[H]^2\right)}$$

*UV-Vis: Changes up on titration – **eqn (36)**:*

$$\Delta A_{obs} = \frac{\varepsilon_{\Delta HG}[G]_0K_1[H] + 2\varepsilon_{\Delta H_2G}[G]_0K_1K_2[H]^2}{1 + K_1[H] + K_1K_2[H]^2}$$

Fluorescence*: Changes in fluorescence up on titration (no dynamic quenching) – **eqn (45)**:*

$$\Delta F_{obs} = \frac{k_{\Delta HG}[G]_0K_1[H] + 2k_{\Delta H_2G}[G]_0K_1K_2[H]^2}{1 + K_1[H] + K_1K_2[H]^2}$$

Calorimetry: Heat of formation upon complex formation:

$$Q = \frac{\Delta H_{HG}V[G]_0K_1[H] + 2\Delta H_{H_2G}V[G]_0K_1K_2[H]^2}{1 + K_1[H] + K_1K_2[H]^2} \quad (48)$$

Chart 3 Summary of equations related to 2 : 1 equilibria.

eqn (17) and (18) would, however, be slightly different assuming again that the guest is "silent" ($Y_G = 0$) as shown in eqn (21) and (22).

3.6 Cooperativity

In the above discussion on 1 : 2 and 2 : 1 systems the relationship of stepwise binding constants K_1 and K_2 and how these are related to K_a for a simple (related) 1 : 1 host–guest system have not been explored. Taking the 1 : 2 system as an example, we start with an empty ditopic host (**H**) with two identical binding sites which we label **A** and **B** (Fig. 3). When the first guest (**G**) binds to this host, it can either bind to site **A**

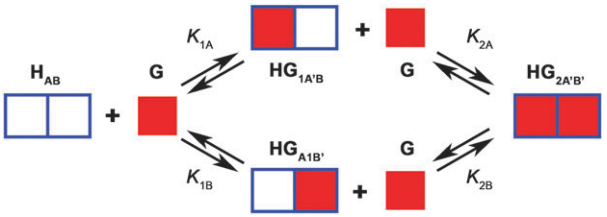

Fig. 3 A schematic explaining the microscopic (K_{1A}, K_{1B}, K_{2A} and K_{2B}) association constants involved in the stepwise formation of a 1 : 2 complex.

to form a $\mathbf{HG_{1A'B}}$ (the prime ′ indicates which site is occupied) complex or to site **B** to form $\mathbf{HG_{1AB'}}$, with the equilibria described by the binding constants K_{1A} and K_{1B}, respectively. When the second **G** binds to the remaining sites in $\mathbf{HG_{1A'B}}$ and $\mathbf{HG_{1AB'}}$, the product in both cases is $\mathbf{HG_{2A'2B'}}$, described by the binding constants K_{2A} and K_{2B} (Fig. 3).

If the binding sites **A** and **B** are truly identical we cannot of course distinguish $\mathbf{HG_{1A'B}}$ and $\mathbf{HG_{1AB'}}$ and hence, determine K_{1A} and K_{1B} (or K_{2A} and K_{2B}) as the physical changes analogous to, *e.g.*, Y_{HG} in eqn (7) for sites **A** and **B** in **HG** are identical. Although the binding constants K_{1A}, K_{1B}, K_{2A} and K_{2B} cannot be measured directly they can be related to the overall stepwise constants K_1 and K_2. Starting with eqn (3), we see that $[HG] = [HG_{1A'B}] + [HG_{1AB'}]$ and hence $K_1 = K_{1A} + K_{1B}$. Likewise, from eqn (4) it is possible to show that $K_2 = K_{2A}K_{2B}/(K_{2A} + K_{2B})$. If the binding sites **A** and **B** are identical, then it follows immediately that $K_{1A} = K_{1B} = K_{1m}$ and $K_{2A} = K_{2B} = K_{2m}$, with K_{1m} and K_{2m} the first and second microscopic binding constants. Although these cannot be measured directly, we can see from the above that $K_1 = 2K_{1m}$ and $K_2 = K_{2m}/2$. Conceptually, the factor of 2 in these relations can be explained on kinetic grounds (see also Section 3.1) as there are two ways for **G** to bind to the empty $\mathbf{H_{AB}}$ so that the observed on-rate (k_1) appears twice as fast while there are two ways for the **G** to come off (k_{-2}) the complex $\mathbf{HG_{2A'B'}}$. If we furthermore assume that there is *no change in the empty remaining site* in $\mathbf{HG_{1A'B}}$ or $\mathbf{HG_{1AB'}}$ and no specific interaction (*e.g.* electrostatic) between two molecules of **G** bound to $\mathbf{HG_{2A'B'}}$, then $K_{1m} = K_{2m}$ which describes classical non-cooperative binding in a 1 : 2 system. We see now that for such non-cooperative binding the stepwise binding constants K_1 and K_2 are related by eqn (25).[10,20]

$$K_1 = 4K_2 \quad (25)$$

This equation is a special case of a more generalised equation formally defining the expected relationship between stepwise binding constants in any non-cooperative system, *i.e.* a system where the binding sites are truly identical and independent of each other.[21] If the binding of the first **G** to $\mathbf{HG_{1A'B}}$ or $\mathbf{HG_{1AB'}}$ does change the binding properties of the remaining site or there is some interaction between the two **G** bound to $\mathbf{HG_{2A'B'}}$(such as electrostatic repulsion), then $K_{1m} \neq K_{2m}$ which describes *cooperative binding*. We can quantify the extent of this cooperativity with the interaction parameter α according to eqn (26).[10,20]

$$\alpha = \frac{4K_2}{K_1} \quad (26)$$

If $\alpha > 1$ the system displays positive cooperativity, if $\alpha < 1$ it displays negative cooperativity and if $\alpha = 1$, we arrive at eqn (25) describing non-cooperative binding.[20] The interaction parameter α can also be used to describe cooperativity (or lack of it) in 2 : 1 binding system; an $\alpha > 1$ would suggest that the formation of a 2 : 1 complex is favourable over the formation of a 1 : 1 complex.[22] When fitting suspected 1 : 2 (or 2 : 1) titration data to quantitative descriptions such as eqn (17) it is good practice to fit the data to both a non-cooperative 1 : 2 model, with K_1 fixed as $4 \times K_2$ (or *vice versa*) as well as cooperative 1 : 2 model where both K_1 and K_2 are optimised simultaneously. Using methods described below, the results should then be compared in order to evaluate if the system of interest does display true cooperative behaviour. There are various methods in the literature for plotting data from cooperative systems, including the above mentioned Scathard and the Hill which is a log–log plot of the fraction of bound [H]/free [H] sites against free [G] that are both popular for the analysis of biomolecular binding.[12,23] These methods have limited value in classical supramolecular titration studies as they require that the free concentration of the guest and host can be measured directly.

Co-operativity is not only observed in these simple *homotropic* (both guests are the same) systems but also in heterotropic (the two guests are not the same) 1 : 2 or 2 : 1 systems as highlighted in Fig. 1b.[6,7] The more complicated definitions of co-operativity that apply to these and other systems such as aggregation (self-assembly) have recently been reviewed[24] and will not be discussed here further.

3.7 More complex systems

Although outside the scope of this *tutorial review*, similar approaches can be used to develop equations for more complex equilibria but not only do these become computationally difficult (solving a quadric or even quintic equations) but also because the increased number of unknown parameters (K_1, K_2, K_3...) makes it difficult to get any meaningful results from the fitting process. Here, simplifications become a necessity. An illustrative example comes from studies on the so-called sandwich 2 : 2 complexes first reported by Anderson.[22] For a more detailed discussion about this and other more complex cyclic system the reader is referred to the excellent work of Ercolani.[25]

4. Determination of stoichiometry

The execution and analysis of a supramolecular titration experiment is heavily dependent on having at least some knowledge of what the stoichiometry ratio, m/n in eqn (2), of host and guest is. Ultimately, one of the main aims of any supramolecular titration experiment will also be to determine the stoichiometry of the system under study with the titration data itself providing a key piece in that puzzle. There are no magical solutions to this challenge but as Connors suggested,[10] a good starting point is to assume a simple 1 : 1 stoichiometry and then look for other evidence to support or contradict that assumption. Connors lists a number of possible methods that can be used including:[10]

(i) The method of continuous variations (Job's method).

(ii) Comparison of stability constants evaluated by different methods,[26] including solubility diagrams.

(iii) Consistency with the host structure and available information on the host–guest complex structure.

(iv) Specific experimental evidence such as isosbestic point(s).

(v) Constancy of stability concentration as the concentration is varied, that is, the success of a stoichiometric model to account for the data.

Of these, the Job's method has gained significant popularity to the point of that researchers have started to ignore looking at other methods to verify stoichiometry even in situations *where the Job's method is not appropriate*. The idea behind it is simple; the concentration of a $\mathbf{H_mG_n}$ ($[H_mG_n]$) complex is at maximum when the [H]/[G] ratio is equal to m/n.[8,11] To do this, the mole fraction (f_G) of the guest is varied while keeping the total concentration of the host and guest constant ($[H]_0 + [G]_0 =$ constant). The concentration of the host–guest complex $[H_mG_n]$ is then plotted against the mole fraction f_G yielding a curve with a maxima at $f_G = n/(m + n)$, which in the case of $m = n$ (*e.g.*, 1 : 1) appears at $f_G = 0.5$.[11] In practice, determining the concentration of $[H_mG_n]$ experimentally may not be straightforward as discussed in Section 3.2. Instead, a property, such as NMR chemical shift or UV-Vis absorption peak, that *seems* to have linear dependence on $[H_mG_n]$ is used and plotted against f_G. For NMR titrations the approach used by Crabtree is typical of this approach as $[H_mG_n]$ is approximated according to eqn (27).[27]

$$[H_mG_n] = \frac{\Delta\delta[H]_0}{\delta_{H_mG_n} - \delta_H} \quad (27)$$

The problem here of course is that only at infinite concentration of $[G]_0$ (when the system is fully saturated) can one obtain an accurate value for δ_{H_mG}, although a reasonable estimate can be obtained by extrapolating the observed δ_{obs} at high $[G]_0/[H]_0$ ratios. Generally speaking though, the Job's method does work well when there is only one type of complex (*e.g.*, 1 : 1) present.[10] When there is more than one complex present, the *Job's method becomes unreliable*.[10,28] This includes many situations with m/n = 1 : 2 or 2 : 1 as these usually include two forms of complexes (*e.g.*, **HG** and $\mathbf{HG_2}$) that have different physical properties, hence the assumption that the physical property of interest (*e.g.*, δ_{obs}) is linearly dependent may not be valid. For similar reasons, the Job's method is likely to fail when either the host or guest aggregates in solution.[9]

When Job's method fails to confirm the simple assumption about 1 : 1 complexation, one must resort to one or preferably more of methods (ii)–(v) above. The first of these, Method (ii) is very useful when supramolecular titrations are combined with other approaches to determine association constants, including kinetic measurements and solubility studies.[26] Solubility studies do though require significant quantities of host (and guest), explaining perhaps why this method has fallen out of favour as the synthetic sophistication of the hosts used in supramolecular chemistry increased in recent times.

Method (iii) is perhaps the simplest but often the most effective of all the approaches available to determine the stoichiometry in host–guest complexes. In modern supramolecular chemistry it is now rare not to have detailed information through X-ray crystallography, 2D-NMR and

Molecular Modelling about the structure of the host and guest and in some cases even the host–guest complex itself. This structural information can make the prediction of stoichiometry quite straightforward and accurate.

Method (iv) relies on specific evidence such as isosbestic points which can be used to confirm that more than one type of complex is present and hence that simple 1 : 1 complexation is not appropriate to describe the system if more than one isosbestic point is observed.[8] The converse is not necessarily true, *i.e.* the absence of more than one isosbestic point cannot be used to rule out more complex stoichiometry such as 1 : 2 complex formation, especially in cases where the cooperative (positive or negative) processes play a significant role.

Connors points out that Method (v) is probably the most generally applicable method for determining stoichiometry. Firstly, if anything other than 1 : 1 stoichiometry is suspected, the data should be fitted to other plausible models (*e.g.*, 1 : 2) and the quality of fit of the different models compared in details, taking into account factors such as the increase in parameters in the fitting process (see also Section 6.2 below). Secondly, and more importantly, it is strongly advisable to carry out the titration at different concentrations and even with different techniques (*e.g.*, NMR and UV-Vis). If a particular model is successful at explaining the data at different concentrations then it can be taken as very strong evidence for that model.

Finally it is worth mentioning the mole ratio method[11] which essentially uses an ordinary binding isotherm from a titration study where the concentration of $[H]_0$ is fixed and the concentration of $[G]_0$ is varied. The apparently linear portions at the beginning and the end of the curve are extrapolated to find a break point, corresponding to the point of the isotherm with the most abrupt change. This break point usually corresponds to the $[G]_0/[H]_0$ stoichiometric ratio and is especially useful for non-cooperative higher-order complexes. This method needs to be applied with care if the system is likely to display significant levels of cooperativity or if the physical changes between the different levels of complexation (*e.g.*, the $\mathbf{H_2G}$ and **HG** complex) are non-linear. Some software packages try also to find the stoichiometry essentially by introducing the n/m ratio as one of the parameters to fit. The n/m is then reported back as a real number which essentially can take on any value between 0 and infinity. This makes sense when studying the binding of small guests to large (bio)macromolecules such as proteins and DNA. Two notes of caution need to be raised here: firstly, most of these software models assume no cooperativity—an assumption that is not always appropriate. Secondly, a non-integer n/m value does of course not describe the real system—*e.g.*, a host will either bind 1 or 2 cations but not 1.3 (there is no host with 0.3 binding sites)! A better approach is to use the results from these programs as a starting point for further analysis of the stoichiometry.

5. Common techniques for supramolecular titrations

5.1 General consideration and preparation of solutions

It is hopefully now clear from the above that a proper understanding of how association constants are determined cannot be gained without looking at the underlying fundamental equilibria and mathematics. In addition, some *chemical understanding of the system* under scrutiny is essential. What sort of association is most likely? Is it probably a case of 1 : 1 host-to-guest interaction or is there multivalency in this system? What influence will this association have on the host's physical properties (UV-Vis, fluorescence, NMR, solubility *etc*...). Will the host aggregate as well? How strong are these associations likely to be? It may seem counter-intuitive to ask such questions as the key purpose of titration experiments in supramolecular chemistry is usually to get proper answers to the above questions! The reality is, however, that having at least some reasonable answer(s) to the questions above will aid enormously in the design of the titration experiments, including what technique will be most appropriate for the titration experiment. In some instances trial titration experiments covering a large range of concentrations may be necessary to probe these questions.

When a supramolecular titration study is carried out one has to first make a decision on what technique is going to be used to follow the physical changes (ΔY) in the system during the course of experiment. The two key concerns here should be:

i. The expected association constant(s).
ii. The expected physical changes (ΔY) upon association.

The expected association constant determines what concentration should be chosen for the host system which in turn will have an influence on the choice of technique. Wilcox,[29] using a parameter defined by Weber as probability of binding or p,[30] showed that it is vital to collect as many data point as possible within the range: $0.2 < p < 0.8$ with p defined according to eqn (28) and (29).

$$p = \frac{[HG]}{[G]_0} \text{ when } [H]_0 \geq [G]_0 \quad (28)$$

$$p = \frac{[HG]}{[H]_0} \text{ when } [H]_0 < [G]_0 \quad (29)$$

Using eqn (13) from Section 3.2, it is possible to calculate p for a range of $[H]_0$, $[L]_0$ and K_a values. When the results are plotted for three fixed $[H]_0$ concentration ranges (10^{-3} M, 10^{-5} M, 10^{-7} M) as a function of K_a and $[L]_0/[H]_0$ (equivalents of guest added) typically employed in NMR, UV-Vis and fluorescence spectroscopy studies, respectively, a revealing pattern appears (Fig. 4) with the shaded areas indicating p in the range of 0.2–0.8 (note that this applies only to 1 : 1 binding systems).

Here it is helpful to introduce the concept of the *dissociation constant*: $K_d = 1/K_a$—the inverse of the association constant K_a. The K_d values are plotted for reference on the right-hand y-axis in Fig. 4 (a thick horizontal line indicated where $K_d = [H]_0$).With this in mind, one can look at three quite different regions in Fig. 4.

If $K_d > [H]_0$ (hence K_a fairly low) then a relatively large excess of $[G]_0$ is required to obtain good p-values. In this situation it would be advisable to collect several data points in the range of 1–50 equivalents of **G** added. Interestingly, if $[H]_0/K_d > 100$, it is not always necessary to have an accurate

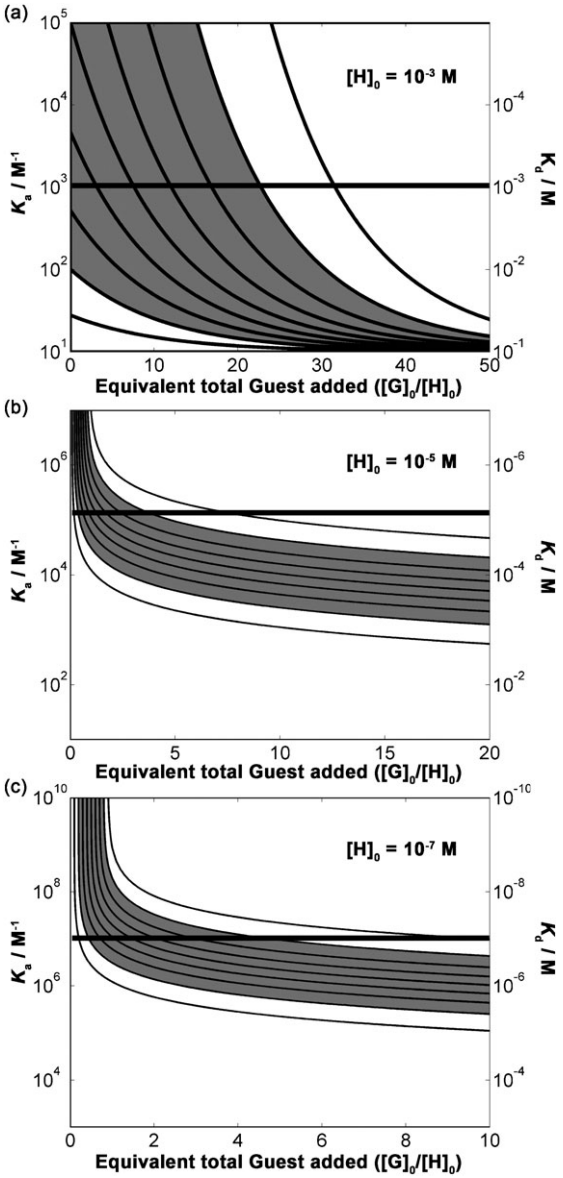

Fig. 4 Probability of binding p according to Weber[30] as a function of equivalent total Guest (G_0) and association constant K_a (left axis) and dissociation constant $K_d = 1/K_a$ (right axis) at different host concentrations. (a) $[H_0] = 10^{-3}$ M; (b) $[H_0] = 10^{-5}$ M; (c) $[H_0] = 10^{-7}$ M. The thick horizontal line indicates where $[H_0] = K_d = 1/K_a$. The contour lines are 0.1 units of p apart with the lowest one shown at $p = 0.1$ and with p in the range of 0.2–0.8 shaded grey. See text for details.

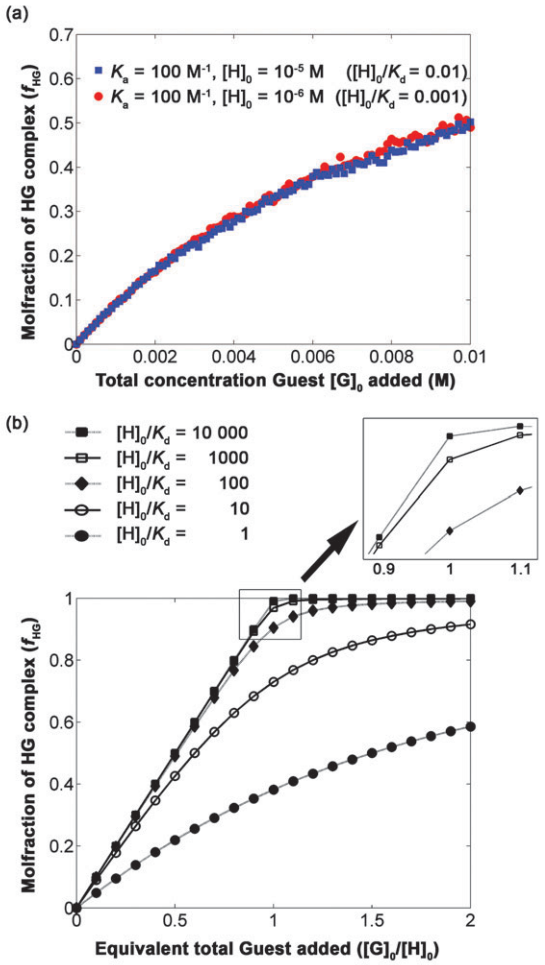

Fig. 5 Illustrative binding isotherms for 1 : 1 complexation scenarios. (a) Simulated binding isotherm (2% added random error to the Y-axis) of two scenarios with $K_a = 100\ M^{-1}$ and $[H]_0 = 10^{-6}$ M (red) and 10^{-5} M (blue). This simulation highlights the fact that at least in the region of $[H]_0/K_d < 0.01$, accurate knowledge of $[H]_0$ is not essential for 1 : 1 complexation. (b) Binding isotherms for different $[H]_0/K_d$ ratio's from 1–10 000. The inset shows the region around 0.9–1.1 equivalents added for $[H]_0/K_d = 1000$–10 000 only. See also text for details.

value for $[H]_0$. For example, if $K_a = 100\ M^{-1}$, the binding isotherms for $[H]_0 = 10^{-5}$ M and $[H]_0 = 10^{-6}$ M become practically indistinguishable (Fig. 5a).

If $K_d < [H]_0$ (hence K_a fairly high) the only data points with good p-values are within the range of $[G]_0 < [H]_0$. In other words, it is essential to obtain as many points as possible between 0–1 equivalents of **G** added.

If $K_d \approx [H]_0$, good p-values are obtained almost anywhere within the range of 0 to >10 equivalent of **G** added. Note that when $K_d = [H]_0 = [G]_0$, then $p = 0.38$.

The fourth scenario to consider is when $K_d \ll [H]_0$, *i.e.*, by a factor of at least 100. Here, it is not enough to look just at the p values obtained. Consider instead what is happening with the non-linear portion of the resulting binding isotherms (Fig. 5b).

Here it becomes clear that once $[H]_0/K_d > 100$, the non-linear portion of the resulting binding isotherms is restricted to a small region around 1 equivalent of guest added. With $[H]_0/K_d > 1000$, it is clear that there is very little "information" content in the isotherms. A Monte Carlo simulation (see below) with 2% added random error to the Y-data highlights the difficulties with getting good data from high $[H]_0/K_d$ ratios. Here, the asymmetrical uncertainty based on the Monte Carlo simulation confidence limits with $[H]_0/K_d = 10$ are *ca.* −7%/+11%, with $[H]_0/K_d = 100$, they are −35%/+75% and with $[H]_0/K_d = 1000$, they are $-67\%/+10^{14}\%$! Even using the global analysis approach discussed below, it is clear from this that reliable results can only be obtained up to a

$[H]_0/K_d = 100$—beyond that only a very crude estimation of the lower limit for K_a can be obtained.

From all the above, it is obvious that with the exception of very weakly bound systems, data points in the range of 0–1.5 equivalents of **G** added are usually the most important data points. There is no fixed rule on how many points should be collected but in the experience of this author, 10 is the bare minimum with 15–20 a desirable target for 1 : 1 binding (especially if the binding is more complex). If there is some prior knowledge on the strength of binding, a good strategy is to collect at least 8–10 points between 0–1.5 equivalents of **G** and then another 10–15 points spaced non-linearly between 1.5–50 equivalents of **G** added. Alternatively, when the strength of binding is completely unknown, one can increase the guest concentration exponentially, *e.g.*, by increasing the concentration of $[G]_0$ by a factor of 3 on every addition, starting at 0.1 equivalents and going up to the highest possible concentration. After about 1–15 addition, the resulting data will cover 3–4 orders of magnitude of the $[G]_0/[H]_0$ ratio and hence complexation should be detected somewhere on that scale even in cases when $K_d > 1000 \times [H]_0$.

In the case of 1 : 2 (and 2 : 1) equilibria the situation is even more complex due to (possible) cooperative effects and differences in the physical properties being measured, that is $Y_{\Delta HG} = Y_{HG} - Y_H$ and $Y_{\Delta HG_2}$ in eqn (17) and (18). As an example let's look a few possible scenarios for a UV-Vis titration where the total host concentration $[H]_0 = 10^{-5}$ M (Fig. 6).

Inspection of Fig. 6 reveals a number of interesting phenomena. Firstly, looking at the distribution of the two different complexes in solution (Fig. 6b), we see that for the statistical binding and positive co-operativity scenarios, the concentration of the 1 : 2 complex **HG_2** only reaches above a mole fraction of 0.2 (roughly corresponding to $p > 0.2$ above) once about 2–3 equivalents of **G** have been added. In the negative co-operativity case, over 10 equivalents are needed to reach a mole fraction of 0.2. In contrast the intermediate 1 : 1 complex **HG** is most pronounced between 1–10 equivalents of **G** except in the case of positive cooperative binding. The latter observation suggests that for highly positive cooperative systems, the detection of the intermediate 1 : 1 **HG** complex can become very difficult.

The situation is further confounded by all the possible combinations that the changes in molar absorptivity $\varepsilon_{\Delta HG}$ and $\varepsilon_{\Delta HG_2}$, can take. In Fig. 6c, the UV-Vis isotherms for three possible scenarios for $\varepsilon_{\Delta HG}$ and $\varepsilon_{\Delta HG_2}$ are plotted for the three different cooperativity scenarios outlined in Fig. 6a. The three $\varepsilon_{\Delta HG}$ and $\varepsilon_{\Delta HG_2}$ scenarios are (i) when $\varepsilon_{\Delta HG}$ and $\varepsilon_{\Delta HG_2}$ are quite similar (30 000 *vs.* 20 000), (ii) when the first is very weak compared to the second ($\varepsilon_{\Delta HG}$ and = 1000, $\varepsilon_{\Delta HG_2}$ = 20 000) and finally (iii) when the second has an opposite almost equal sign to the first ($\varepsilon_{\Delta HG}$ = 30 000 and $\varepsilon_{\Delta HG_2} = -20\,000$).

This simulation shows that the differences in molar absorptivity can have at least the same if not more effect on the resulting binding isotherm than changes in cooperativity. Further, unless the signs of the change in molar absorptivity are opposite, the differences between the binding isotherms are not that large, especially in the early stages of the titration.

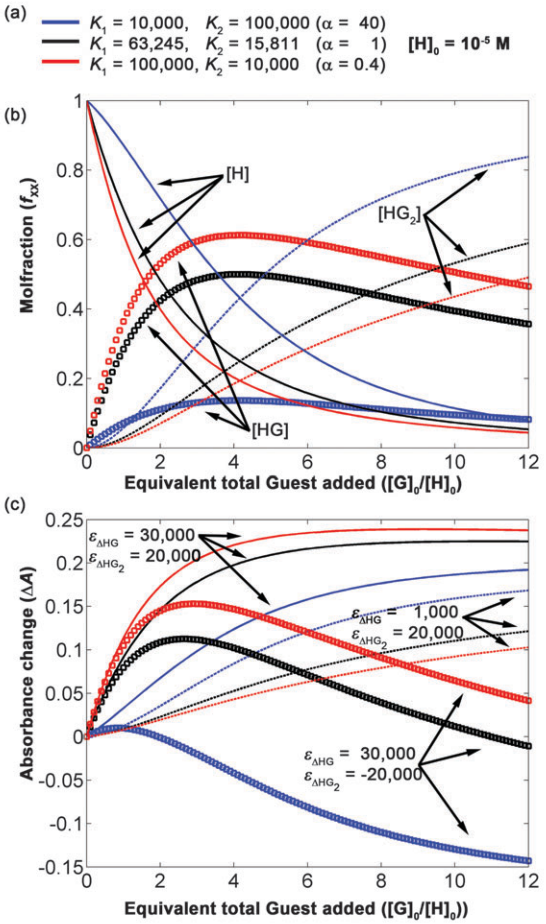

Fig. 6 Simulated UV-Vis titrations for a 1 : 2 equilibria with $[H]_0 = 10^{-5}$ M and $\beta_{12} = K_1K_2 = 10^9$ M^{-2}. (a) Colour coding for the three different scenarios: blue = positive co-operativity ($\alpha = 40$), black = statistical binding ($\alpha = 1$) and red = negative co-operativity ($\alpha = 0.025$). (b) Mole fraction distribution of [H], [HG] and [HG_2] (see arrows). (c) The resulting UV-Vis binding isotherms depending on three possible scenarios for $\varepsilon_{\Delta HG}$ and $\varepsilon_{\Delta HG_2}$ (see arrows). Note that two of the binding isotherms for the negative co-operativity ($\alpha = 0.025$) scenario coincide. See also text for details.

The take-home message from Fig. 6 is that one needs to obtain as many data points from the whole range of at least 0–10 equivalent of **G** added as possible, with special emphasis on the first part of that range.

Finally, having decided on the concentration range and the number of addition (titration) points to use, it is worth considering a few tips on how to prepare solutions for supramolecular titrations. It goes without saying that the purity of the host and guest is of utmost importance, especially with optical techniques (UV-Vis, fluorescence *etc.*) as small impurities (which may not be detectable by NMR) can cause peaks to appear (or disappear) unrelated to complex formation. The concentration of the host should also be kept constant throughout the titration to avoid complications arising from (minor) aggregation and the application of complicating dilution factor corrections in the quantitative equations used to describe the system of interest.

The easiest way to do this is to make up the guest solution from the same host stock solution used in the titration. In the case of an NMR titration one could therefore make up a 2 mL solution of the host at concentration $[H]_0$ and then use 1 mL of that solution to make up the guest solution at *ca.* 20–100 fold higher concentration ($[G]_0$ = 20–100$[H]_0$). The remaining host solution (*ca.* 600–1000 μL) is then transferred to an NMR tube. Note that all these volumes need to be measured accurately which is preferably done with an accurate microbalance (and then convert to volumes using the density of the solvent) rather than relying on volumetric flasks or pipettes. The guest solution is then added in small quantities (1–10 μL) to the NMR tube using an accurate microsyringe (such as Hamilton® syringes for organics or Eppendorf® pipettes for aqueous solutions).

5.2 NMR titrations

The most informative technique in most situations is ^{1}H NMR. Other forms (^{13}C, ^{19}F *etc.*) of NMR are also applicable. Apart from the quantitative information that an NMR titration can yield, the relative shifts and changes in symmetry can often give valuable information about how the host and guest(s) are interacting and the stoichiometry of interaction. This information can be of significant benefit even in situations where complete quantitative data cannot be obtained from the NMR titration as shown in our work on the binding of a ditopic porphyrin host to viologen where symmetry changes were indicative of 1 : 1 and 1 : 2 complex formation.[31]

Classical approaches for data analysis of NMR titrations assume that the resonance (δ) of interest is the weighted average of the free host (**H**) and the bound host in the complex (**HG**) in the experiment according to eqn (30) in Chart 1 in the case of a simple 1 : 1 system.[11] We immediately recognise this as a special case of eqn (13) above in Section 3.2. We can proceed directly from the relevant equations in Sections 3.4 and 3.5 to the descriptive quantitative expressions that outline the expected changes in chemical shifts ($\Delta\delta$) assuming a 1 : 2 or 2 : 1 complexation, according to eqn (31) or (32), respectively (Chart 2).

With modern NMR instruments it is possible to obtain good quality spectra with sub-millimolar concentrations (routinely now as low as 10^{-4} M), suggesting from the above discussion that NMR might be suitable for K_a up to and even above 10^6 M^{-1}. That said, many literature references will state that 10^5 M^{-1} is the limit for NMR titration experiments.[29]

Here, one has also to take into account the relative exchange rates within the host–guest or in other words, the previously mentioned (Section 3.1) relationship between the equilibrium association constant and the kinetics on/off rates ($K_a = k_1/k_{-1}$) and the timescale of the NMR experiment. The real limiting factor for NMR titrations is therefore whether the system of interest is in the fast or slow exchange region under the conditions used. Besides the association constants and the related on/off rate constants, the magnitude of the observed complexation induced chemical shift ($\Delta\delta$) is related to the mole fraction of the free host divided by the off rates (f_H/k_{-1}) or equivalently, the combined inverse of the lifetimes of nucleus of the complex ($1/\tau_{HG}$) and the free host ($1/\tau_H$). It has been shown that the simple assumptions behind eqn (30) and related equations break down if the system is not in the fast exchange region as defined by eqn (33).[32]

$$2\pi|\nu_H - \nu_{HG}| \ll \left(\frac{1}{\tau_H} + \frac{1}{\tau_{HG}}\right) \qquad (33)$$

Inspection of eqn (33) shows that it can actually be beneficial to focus on resonances with relatively small complexation shift ($\Delta\delta$) when analysing titration curves as these might still be in the fast exchange region while those with larger $\Delta\delta$ may have broadened out due to slow exchange! That said, K_a = 10^5 M^{-1} remains the usual practical limit for direct NMR titrations in the fast exchange region (competition experiments being one notable exception, *e.g.*, the work of Wilcox *et al.*).[33] It may be tempting to think that in the (very) slow exchange region of NMR, one could obtain an association constant directly from the relative ratios of the free and bound host, however, can be difficult in practice due to complications that arise in the intermediate-to-slow region with the size (amplitude) of the observed resonances[32] and the usual limitation of obtaining accurate (quantitative) integration from NMR experiments.

5.3 UV-Vis spectroscopy

The second most common method for supramolecular titration experiment is probably UV-Vis spectroscopy. With the right chromophore (*e.g.*, in the case of porphyrins), host concentration in the sub-micromolar (10^{-7} M) can be applied, making the determination of association constants as high as 10^9 M^{-1} in simple 1 : 1 systems possible (albeit difficult) with $K_d/[H]_0$ = 100 as discussed above. The concentrations chosen must lie within the region where the absorption peak(s) of interest in both the host and its complex are within the limits of the Beer–Lamberts Law ($A = bc\varepsilon$, with $A < 1$). Additionally, it is desirable that the guest added does not have any absorption in the region of interest as this simplifies the system considerably—fortunately, this is usually the case in simple supramolecular systems (*e.g.*, upon addition of simple cations or anions to a host). It is also important that the complexation causes a notable change in the UV-Vis spectra as the usual approach for analysis of UV-Vis titration data assumes a significant change in the molar absorptivity (ε) upon complexation according to eqn (34) (Chart 1) which is a special case of eqn (15) from Section 3.2 above. Similarly, with 1 : 2 and 2 : 1 binding, one obtains eqn (35) (Chart 2) and eqn (36) (Chart 3) that are derived from eqn (18) and (21).

Although applicable to all the different methods available (including NMR), titration by UV-Vis spectroscopy is particularly vulnerable to dilution and temperature effects (all supramolecular titration experiments need some temperature control) and the presence of impurities in either host or guest solutions. If a low concentration of the host is required, special care needs to be taken in weighing out samples and solutions so that the concentration of the host can still be determined with good accuracy. In certain situations (*e.g.*, when only minute quantities of host are available) it is not possible to accurately determine the concentration of the

host and the host concentration must then be included as one of the unknown parameters in the fitting process. It should also be noted again that when $[H]_0/K_d > 100$, it is not necessary to have an accurate value for $[H]_0$ in the case of 1 : 1 binding (Fig. 5a).

5.4 Fluorescence spectroscopy

The third most popular and perhaps the most sensitive technique is fluorescence (and other related luminescence) spectroscopy. The phenomenal sensitivity of this technique makes routine measurements in the sub-micromolar, even nanomolar (nM) range possible and hence, fluorescence spectroscopy is ideal for the determination of very large association constants ($K_a > 10^6$ M^{-1}). In fact, fluorescence titration must be carried out at relatively low concentration or ideally where the absorbance at the excitation wavelength used is less than 0.05 ($A < 0.05$). Above these concentrations the fluorescence response (F) is no longer linearly dependent on the light absorbed. Additionally, inner filter effects will also cause deviations from linearity at higher concentration.[34] The observed fluorescence at low concentration of a species X can usually be described by eqn (37).[34]

$$F = I_0\Phi\varepsilon b[X] = k_X[X] \quad (37)$$

Fluorescence is a particularly useful technique in the case when only one of the species in solution is fluorescently active, *i.e.* when either the free host or guest is fluorescent "silent" or inactive and the fluorescence of the remaining species is either turned "off" (quenched) or "on" upon complexation. If quenching plays a role, it is necessary to differentiate between static and dynamic (collisional) quenching, with only the former of real significance for supramolecular binding studies. Dynamic quenching is usually measured by plotting the ratio of the initial (F_0) and measured (F) fluorescence intensity ratio (F_0/F) against the concentration of the quencher [Q] according to the Stern–Volmer relation $F_0/F = 1 + K_{SV}[Q]$, with K_{SV} = the Stern–Volmer constant. Unfortunately, pure 1 : 1 static quenching follows a nearly identical relation: $F_0/F = 1 + K_a[Q]$, with [Q] = the *free concentration* of the quencher (guest) and K_a is of course the familiar association constant of interest in supramolecular binding studies. In many cases the observed quenching is a mixture of both static and dynamic quenching which can lead to some complication in the analysis of the titration data. With this in mind, there are three general scenarios that need to be taken into account when analysing data from supramolecular fluorescence titrations (we will skip here the derivation for the 1 : 2 and 2 : 1 systems—these can be derived from the generic relations in Section 3—see Charts 2 and 3).

The most simple case is the situation when only the complex formed is fluorescent (complexation turns fluorescence on) with the observed fluorescence described by eqn (38).

$$F = k_{HG}[HG] \quad (38)$$

The next scenario is when both the free host and the complex are fluorescent. Before any guest is added to the solution the observed fluorescence can be described by eqn (39).[10]

$$F_0 = k_H^0[H]_0 \quad (39)$$

After the addition of a guest, the resulting fluorescence will follow eqn (40).[10]

$$F = k_H[H] + k_{HG}[HG] \quad (40)$$

It is important to recognise that k_H^0 may not be equal to k_H, essentially due to complications arising from dynamic quenching of the free host. Many researchers find it convenient to combine eqn (39) and (40) with the aid of eqn (9) to get eqn (41).[10]

$$\frac{F}{F_0} = \frac{k_H/k_H^0 + (k_{HG}/k_H^0)K_a[G]}{1 + K_a[G]} \quad (41)$$

In the special situation where the complex is fluorescently silent (quenched) and $k_H^0 = k_H$ (no dynamic quenching) we obtain an equation that looks identical to the Stern–Volmer equation describing classical static quenching according to eqn (42).[10]

$$\frac{F_0}{F} = 1 + K_a[G] \quad (42)$$

Likewise, if $k_H{}^0 = k_H$ (no dynamic quenching of the host) but the complex is still fluorescently active ($k_{HG} \neq 0$) we can use the generic approach from Section 3.2 with eqn (15), to obtain eqn (43).

$$\Delta F_{obs} = k_{\Delta HG}([HG]) \quad (43)$$

With $\Delta F_{obs} = F_{obs} - F_0$, the observed fluorescence change and $k_{\Delta HG} = k_{HG} - k_H$, the change in the proportionality constant between the complex and the free host.

Obviously, eqn (43) is almost identical to eqn (34) for UV-Vis titration and provided that the assumption $k_H^0 = k_H$ is valid one can also derive equations similar to eqn (35) and (36) to describe 1 : 2 and 2 : 1 binding for fluorescence titrations according to eqn (44) (Chart 2) and eqn (45) (Chart 3).

5.5 Other methods and summary of techniques for supramolecular titrations

The three examples above in combination with the general equations in Section 3 can be used to develop expressions for almost any supramolecular titration techniques imaginable. Calorimetry represents a powerful example of other methods used in supramolecular titration. This technique relies on measuring enthalpy (H) increases on the addition of a guest to a host in a specially designed apparatus measuring the heat (Q) formed or absorbed (usually a isothermal calorimeter—ITC). After taking care of eliminating dilution effects (see also above in Section 5.1), the heat measured is simply related to the molar enthalpy (ΔH) multiplied by the number of moles of the complex formed which can be obtained from the molar concentration of the complex ([HG]) and the volume of solution (V). Hence in the case of a simple 1 : 1 equilibria we again use eqn (15) from Section 3.2 above to obtain eqn (46) (Chart 1). Similarly, with 1 : 2 and 2 : 1 binding

one obtains eqn (47) and (48) that are derived from eqn (18) and (21).

The most powerful feature of calorimetric titrations is that not only does it yield the free energy (ΔG) changes *via* the association constant according to eqn (49) but also the enthalpy and thus the entropy (ΔS) change can also be obtained from eqn (50).

$$\Delta G = -RT\ln(K) \quad (49)$$

$$\Delta G = \Delta H - T\Delta S \quad (50)$$

6. Data analysis and interpretation

6.1 Fitting data: choosing a model, software considerations and global analysis

Having done all the hard work to obtain the supramolecular titration one might hope that analysing it was as simple as pressing a button on a computer. However, this is not the case, even with the most sophisticated data analysis software. Many chemists would be familiar with a similar situation from the field of computational chemistry; with modern modelling software it might look easy to build a molecule and optimise its structure but the fact still is that unless you understand reasonably well the way you enter the initial model, set up the simulation, which force field or DFT basis set you use and so on, the final results may not be worth all that much. The parallels don't end there: in the past most people in both fields had to write their own software code whereas now commercially available software packages appear to make the task more straightforward. This still doesn't exclude researchers from trying to understand what the key steps are.

Briefly, the process could be divided into: choosing a model, making adjustment to the model, transforming data (if necessary), weighing data, deciding on whether to use global analysis, choosing a minimisation algorithm, initial guesses of parameters, iteration, displaying and plotting results and analysing the quality of the fit obtained. Covering all of these aspects of data analysis is beyond the scope of this *tutorial review*, but here the focus will be a few key aspects of this process including: choosing and making adjustments to model, some software considerations and global analysis. The following section will the discuss how to analyse the quality of fit and estimation of uncertainties. For more details and other aspects of data fitting, interested readers are encouraged to read the work of Motulsky.[18,19]

In Section 4 it was emphasised that sometimes the best approach involves actually trying two or more approaches. The case of suspected 1 : 2 equilibria is a classic example. Comparing eqn (13) and (34) (Chart 1) with eqn (16) and (36) (Chart 2) shows that we have only 2 parameters to fit UV-Vis data to a 1 : 1 model (K_a and $\varepsilon_{\Delta HG}$) while there are up to four in the case of a 1 : 2 model (K_1, K_2, $\varepsilon_{\Delta HG}$ and $\varepsilon_{\Delta HG_2}$). However, if we assume that the 1 : 2 binding is statistical, we can fix $K_2 = 4K_1$, reducing the number of parameters by one. Naturally, the model with the highest number of parameters (cooperative 1 : 2 binding) should give us the best fit. What we must do is to fit the data to all three models (1 : 2 cooperative, 1 : 2 statistical and 1 : 1 binding) and then compare the overall quality of fit. Only if going from a simple (1 : 1) to a more complex (*e.g.*, cooperative 1 : 2) model gives truly significant improvements in fitting can we justify our choice of that model.

Choosing and sometimes modifying a model is the first crucial step. The key issue to be aware of is that the more parameters you fit, the better the fit will be. Almost any data could be fitted to a polynomial equation ($y = a + bx + cx^2 + dx^3 + \cdots$) if it is expanded to high enough *n*th-order. This doesn't mean that it has any physical meaning. On the other hand there is sometimes a valid case for introducing extra parameters, for instance, to correct a baseline[19] or as mentioned, to estimate the concentration of the host $[H]_0$ when its concentration and purity are not well established (*e.g.*, in dilute solutions for UV-Vis).

When it comes to the actual data fitting, there are essentially two different types of software available: (i) commercial software packages dedicated to fitting data to binding models such as GraphPad®[19] and custom written programs that often run as macros or routines with generic data and mathematical programs such as Excel®, Matlab® or Mathematica®. Commercially available software packages have the advantage of being relatively easy to use in terms of loading the data with an output that usually includes additional information such as residual plots and estimation of uncertainties (See Section 6.2 below). Their main drawbacks usually are lack of information about the underlying equations used in the fitting and data analysis process. In addition, these packages are usually targeted to customers in biochemistry or pharmacology, which sometimes results in NMR titrations being left out. Custom written packages on the other hand are usually written by people who are active in the area of supramolecular chemistry and hence well aligned with the methods discussed here. The disadvantages of these packages are that they may not have been verified thoroughly for errors and mistakes. Additionally, they may not be particularly user friendly, requiring a fairly high level of software familiarity or even programming expertise of the user.

Although there is frequently a lot of discussion on the relative merits of the different minimisation algorithms used in these software packages, such as the Gauss–Newton, Marquardt and Simplex,[18] the reality is that other features are usually much more important. To start with, does the software handle the cubic equations (16) and (20) properly? These equations can be solved directly in some programs (*e.g.*, Mathematica and Matlab) but other programs may employ the method of successive approximation to solve these equations. The latter approach fails more frequently which may explain that several software packages do not handle 1 : 2 or 2 : 1 binding models at all (see though Section 6.2). As we will see below, global analysis is another method that is particularly useful for dealing with more complex binding equilibria but this feature is also lacking in many software packages. An example of a custom written package (for Matlab) that does provide both a direct solution to eqn (16) and (20) and global analysis options is provided in the ESI† of this paper by the author.

Quite often, the titration yields more than one set of data, for instance in ^{1}H NMR titration where a noticeable shift is

often observed for more than one proton resonance upon complexation of the host. The same applies to UV-Vis data—the formation of a complex often leads to one increase at one particular wavelength while another one is decreasing. The global analysis method is ideal for analysing this sort of data. The idea is simple: take a family of data sets that share at least one parameter and fit them simultaneously. For instance, if i noticeable shifts in proton resonances up on a ^{1}H NMR titration are observed, we modify or expand eqn (30) to obtain eqn (51).

$$\sum_{j=1}^{n} \Delta\delta_i = \sum_{j=1}^{n} \delta_{\Delta HG_i} \left(\frac{[HG]}{[H]_0} \right) \qquad (51)$$

Why would we want to do this, especially just after noting that increasing the number of parameters can artificially make the fitting process better? Here we need to note that what we are actually doing is to double the number of data points we are fitting. But this only explains part of the story as outlined in an excellent paper by Beechem.[35] In simple terms, what *global analysis* really does is to tighten the so-called error surface considerably compared to single-experiment or *local fitting* and this more than compensates for the increase in the number of parameters. The technique is exceptionally useful in the analysis of multiple equilibria (*e.g.*, 1 : 2 complex formation).

The benefits of global *vs.* local fitting of data are best explained by an example. Here two of the outcomes in Fig. 6 are considered, both having the same binding constants but different changes in $\varepsilon_{\Delta HG}$ and $\varepsilon_{\Delta HG_2}$ for these data sets. A little ($\pm$1% standard deviation) random noise is added to the simulated isotherms and the resulting data fitted as two single experiments (local fitting) and also globally in line with eqn (51) as shown in Fig. 7.

The results in Fig. 7 reveal a number of characteristics of global over other fitting procedures. Firstly, the results from the global fitting are much closer to the "true" values (the one used at the start of this simulation) than either of the two local fits. Secondly the asymptoting standard error (see cautions about this type of uncertainty estimation below) is quite acceptable, *i.e.*, less than 20% for all parameters compared to up to 40%–140% for the two local fits. Thirdly, contrary to what one might expect, the parameters obtained from the global fits are not just simply the average of the local fits (*cf.*, K_1). Finally, it is noteworthy that this does not happen because the overall fit is very different—the isotherms obtained from the local *vs.* global fit are very similar (Fig. 7b inset). A good indicator of the quality of the fit is the standard error of the y estimate—SE_y—which is defined by eqn (52) and related to chi-squared (χ^2) in eqn (53).[36]

$$SE_y = \sqrt{\frac{\sum (y_{data} - y_{calc})^2}{N - k}} \qquad (52)$$

$$\chi^2 = \sqrt{\frac{\sum (y_{data} - y_{calc})^2}{N - k - 1}} \qquad (53)$$

Interestingly, the SE_y is actually not any better (lower) for the global fit than the better two of the local fits. We also see from

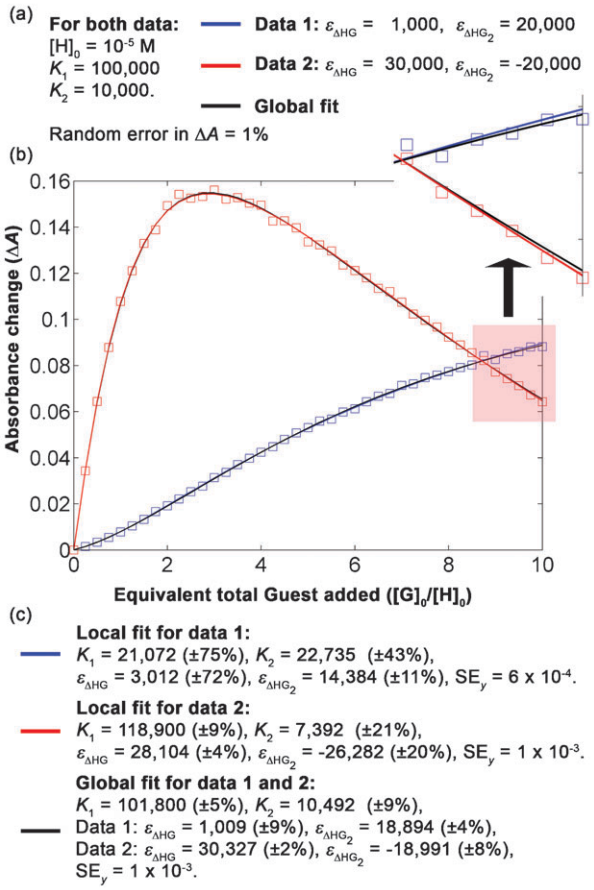

Fig. 7 Local *vs.* global fit for a hypothetical UV-Vis titration for a 1 : 2 complexation equilibria. (a) The initial conditions (see also Fig. 6), the colour coding for the two different data and the global fit. The data points (red and blue squares) are generated from eqn (35) using the parameters shown here with an additional random noise of $\pm$1%. (b) The resulting binding isotherms with the local and global fits (solid lines) superimposed. The enlarged area (red shaded square) shows the similarities of the local and global fitted isotherms. (c) The results from the local and global fitting process with the asymptoting standard error (see Section 6.3) in parentheses. SE_y = standard error of the y estimate—see eqn (52). See also text for details.

this that having a small change in one of the physical properties ($\varepsilon_{\Delta HG}$), in data set 1 is largely responsible for the poor results obtained from that local fit. This is because the fitting processing can alter $\varepsilon_{\Delta HG}$ almost at will while also varying K_1 until a good apparent fit is obtained. The global fitting process on the other hand has no problems in obtaining a good estimate for the small change in $\varepsilon_{\Delta HG}$ while at the same time uses the underlying data to narrow the error-surface around K_1 and the other parameters in this system. Global analysis is in fact particularly powerful in dealing with complex situations such as cooperative 1 : 2 binding systems.

6.2 Evaluating the quality of the fit and estimating uncertainty

The example in Fig. 7 also illustrates the importance of analysing the quality of the fit obtained. This is not a trivial matter and there are no rigorous procedures (bullet proof) to follow here. There are, however, a few basic tests and

strategies that can be applied here.[19] The first one is a visual examination of the fit or more importantly, the residual plot. This is a plot of y_{fit}, fit for each data point, *i.e.*, the difference between measured y_{data} and calculated y_{calc} in eqn (52), as a function of the total guest ($[G_0]$) added. Typically y_{fit} is also normalised by dividing the y_{fit} with SE_y in eqn (52). Ideally, the residual plot should be randomly distributed above and below zero.[37] A systematic trend in the residual plot indicates a poor fit. The residual plot also helps identifying "poor" data points,[37] which may need to be rejected in further analysis, especially if magnitude of y_{fit} is greater than $3 \times SE_y$.[38]

In some situations, physically non-sensible parameters (very low or very high ΔY, negative K_a *etc.*) will suggest that something is wrong with the model. Here, the researcher will have to use their background knowledge; a $k_H = 10^6$ for a fluorescence titration experiment might make sense while $\delta_{\Delta HG} = 100$ ppm for an ^1H NMR titration doesn't make sense. In extreme situations, the data simply don't fit the model with the software used returning answers such as "didn't converge", "Hessian matrix is unstable", "can't divide by zero". This is in some cases a clear sign that something is wrong with the program although it is wise to check the fitting procedure again with different initial guesses.[19] Running a fitting program again with slightly different initial guesses is in any case a good test of how "stable" or close to the real global minimum the fitted solution is. In the case of 1 : 2 binding and more complex situation it is also possible that the program used doesn't handle the cubic eqn (16) and (20) leading to a "false negative" results.

If the data fits reasonably to more than one possible binding model (*cf.* Section 4), one must try to compare the quality of the fits in a quantitative fashion. This is quite often non-trivial and there are many examples where no simple statistical methods can be applied for this task.[19] A natural starting point for this might be to compare the sum of squared fits—$ss_{yy} = \Sigma(y_{data} - y_{calc})^2$—the parameters usually minimized in the fitting process. The SE_y introduced in eqn (52) is though probably a better tool as it tries to make some corrections for the number of parameters used. Another approach that this author has used is the covariance of the fit,[7,9,31] cov_{fit}, defined by eqn (54) as the (co)variance of the residuals ($y_{data} - y_{calc}$) divided by the (co)variance of the raw data.

$$cov_{fit} = \frac{cov(y_{fit})}{cov(y_{data})} \quad (54)$$

The cov_{fit} gives a numerical representation of how random is the scatter in the residual plot relative to the scatter in the original data. This quantity is usually relatively insensitive to the number of parameters used in the fitting process.

Regardless of what quantity (ss_{yy}, SE_y, χ^2, cov_{fit} *etc.*) is used, when comparing two models the one that has a larger number of adjustable parameters (*e.g.*, statistical *vs.* cooperative 1 : 2 equilibria) will usually seem better (lower). This "improvement" in the fit when increasing the number of parameters can be misleading. Here it is suggested that only if there is a significant (*e.g.*, >2–3 fold) improvement in, say cov_{fit}, when comparing two models, should one select the more complex model. Other tests, such as *F*-test can be used provided the models are *nested*, *i.e.*, one of the models is a simpler case of another.[19]

The final, but perhaps most important step in the data analysis is the estimation of the uncertainty. Here one might be tempted to ask if we need to estimate the measurement uncertainty. The answer is of course yes as it confirms how reliable the results are—indeed it has been stated that any analysis without proper information of the reliability is useless![38]

One of the key downsides of non-linear regression analysis is that there are no straightforward methods for estimating the uncertainty of the results obtained from it. The "standard error" reported by most software packages is the so-called asymptotic standard error.[19] This standard error is then used to calculate 95% confidence intervals (which can be approximated as the 2 × the standard error). This error is only "approximate" and it also relies on a few key assumptions regarding the non-linear regression process, including that the error is only on the physical property (Y) measured. This is rarely true as there will always be some (or quite large) uncertainties on the concentrations of the host $[H]_0$ and guest $[G]_0$—the quantities that form the x-axis in the fitting process (*cf.* Fig. 7).

In addition, the very method used to calculate the asymptotic errors introduces assumptions that are usually not valid for a supramolecular system. The details are beyond the scope of this tutorial review but a key step here involves transforming the regression problem into a matrix equation and then a matrix inversion resulting in the so-called variance–covariance matrix. The final results of this assumes that the confidence interval is symmetrical (+/− are the same) around the parameter of interest,[19] which is probably not true in most cases. These factors generally result in underestimation of the realistic uncertainty. Additionally, matrix inversion requires division with the determinant of the original matrix and sometimes this determinant is a very small number, resulting in unstable or even a singular matrix (if determinant is zero).[8] The resulting asymptotic error will therefore be meaningless. This problem is particularly common with more elaborate equilibria such as 1 : 2 equilibria where the underlying matrix manipulations are very complex.

A more rigorous but more intensive approach involves some form of grid search such as the weakened grid search method proposed by Beechem.[35] Here, one parameter is fixed at different values while all others are optimised in the usual manner. The resulting ss_{yy} (or χ^2) is then plotted as a function of the different values for the fixed parameters and the *F*-statistics test is used to determine the increase in ss_{yy} that is significant at a predetermined confidence interval (*e.g.*, 67%) for that parameter. The procedure is then repeated for the other parameters yielding a cross-section of the error surface for all these parameters of interest. In the full grid search method all possible combination of the parameters are varied at certain intervals to map out the error surface and obtain the confidence intervals *via* *F*-statistics, but this method is quite computationally intensive. Even more intensive but perhaps most rigorous of these is a Monte Carlo approach where a random scatter is added to the "ideal data" generated from the parameters obtained (akin to the isotherms in Fig. 6).

We then add random scatter to this data and optimise again (as in Fig. 7) and then repeat the procedure *n*-times to obtain a statistically meaningful distribution of the fitted parameters. Taking Data 2 in Fig. 7 as an example, this yields unsymmetrical and generally a little larger confidence limits for the parameters of interest than the one obtained from the asymptotic error (*e.g.*, for K_1 the limits are 97 000–140 000 with the asymptotic method but 95 000–150 000 with the Monte Carlo approach). The advantage of this approach is that scatter can also be applied to *x*-axis data such as the concentrations of the host $[H]_0$ and guest $[G]_0$. The disadvantages are the computational intensity required and the subjective element involved in deciding the magnitude of the scatter added to the idealised or raw data.

Perhaps the most straightforward method to estimate the uncertainty is simply to carry out the titration several times. Although at least 6–8 repeats are required to obtain truly statistically significant results,[38] a reasonable estimate of the spread in results can usually be obtained after repeating the titration three times and multiplying the resulting standard deviation by two to obtain an approximation of the 95% confidence interval. Repeating the experiment and fitting it again is also a good test of the robustness of the model (*e.g.*, 1 : 2 *vs.* 1 : 1) chosen.

Finally it is worth pointing out that differences and uncertainties in association constants should also be considered in terms of the free energy changes (ΔG) involved. As mentioned in Section 2, doubling an association constant increases ΔG by—1.7 kJ mol^{-1}—equating to 18% increase in energy if the initial K_a = 50 M^{-1} but only 5% if it was K_a = 500 000 M^{-1}. Taking this into account, it follows that a measurement uncertainty of 20% on K_a when K_a is $<$1000 M^{-1} is much more significant than if K_a $>$ 100 000 M^{-1}. Likewise, it is a good practice to compare the stepwise equilibria in cooperative 1 : 2 or 2 : 1 equilibria in terms of the free energy changes in each step according to eqn (55).[31]

$$\Delta G_X = -RT\ln K_{X_i} \quad (55)$$

For a simple 1 : 2 or 2 : 1 equilibria, $K_{1i} = K_1/2$ and $K_{2i} = 2K_2$.[8,31] Consequently, apparent differences in binding constants for a negative cooperative and a positive cooperative binding 1 : 2 equilibria look considerably different when analysed in terms of stepwise free energy changes or $\Delta\Delta G = \Delta G_2 - \Delta G_1$.[31] For instance if K_1 = 10 000 M^{-1} and K_2 = 1000 M^{-1} then $\Delta\Delta G$ = 2.2 kJ mol^{-1} whereas if K_1 = 1000 M^{-1} and K_2 = 10 000 M^{-1} then it follows that $\Delta\Delta G$ = −9.1 kJ mol^{-1}.

7. Conclusions

This *tutorial review* has highlighted the key issues that have to be addressed when carrying out and analysing data from supramolecular titration experiments. The key message is that knowledge is power and one should approach these problems with a critical and curious mindset. As in any form of quantitative or analytic chemistry, the starting point is a systematic approach to minimising external uncertainties and errors, such as impurities and concentration variations. Modern computer software and hardware improvements now allow researchers to use powerful tools such as Global Analysis and Monte Carlo simulations. Importantly, 1 : 2 and related equilibria can now be accurately dealt with by directly solving the underlying key cubic equation. Other more complex systems can also be tackled using modern mathematical and data analysis software packages and powerful search algorithms. This includes the use of genetic search algorithms which are particularly useful in tackling problems with several local minima surrounding the desired global minima. Lastly, blind data analysis will never be a substitute for our chemical intuition and knowledge, *e.g.*, when tackling the important question of stoichiometry. Our supramolecular knowledge base is of course heavily based on prior information from quantitative analysis—this chemistry knowledge is likely to continue to expand rapidly as the field matures.

Appendix A—abbreviations and symbols used in this review

H	host;
G	guest;
X	species X;
$\mathbf{H_xG_x}$	host–guest complex of interest;
HG	1 : 1 complex;
$\mathbf{HG_2}$	1 : 2 complex;
$\mathbf{H_2G}$	2 : 1 complex;
$[H_0]$	the total (or initial) concentration of the host;
$[G_0]$	the total (or initial) concentration of the guest;
[H]	molar concentration of the free (unbound) host;
[G]	molar concentration of the free (unbound) guest;
[X]	molar concentration of species **X**;
[HG]	molar concentration of the 1 : 1 complex;
$[HG_2]$	molar concentration of the 1 : 2 complex;
$[H_2G]$	molar concentration of the 2 : 1 complex;
K_x	the thermodynamic association constant for equilibria x;
k_1	forward rate constant;
k_{-1}	backward rate constant;
K_a	k_1/k_{-1}, association constant;
K_d	$1/K_a$, dissociation constant;
β_{mn}	overall association constant for an *m* : *n* host–guest complex;
K_1	the first stepwise association constant;
K_2	the second stepwise association constant;
K_{11}	K_1 for 1 : 2 equilibria in some literature;
K_{12}	K_2 for 1 : 2 equilibria in some literature;
K_{21}	K_1 for 2 : 1 equilibria in some literature;
K_{22}	K_2 for 2 : 1 equilibria in some literature;
K_{X_i}	the microscopic K_X for site *i* in a mutlivalent host;
K_{1A}	the microscopic K_1 for site A in a ditopic host;
K_{1B}	the microscopic K_1 for site B in a ditopic host;
K_{2A}	the microscopic K_2 for site A in a ditopic host;
K_{2B}	the microscopic K_2 for site B in a ditopic host;
K_{1A}	K_1 for site A in a ditopic host (1 : 2 equilibria);
K_{1B}	K_1 for site B in a ditopic host (1 : 2 equilibria);
ΔG_x	the free energy change for step *x* in multistep equilibria;
$\Delta\Delta G$	$\Delta G_2 - \Delta G_1$;
f_H	mole fraction of the host **H**;

f_G mole fraction of the guest **G**;
f_X mole fraction of species **X**;
f_{HG} mole fraction of the complex **HG**;
f_{HG_2} mole fraction of the complex **HG$_2$**;
f_{H_2G} mole fraction of the complex **H$_2$G**;
Y the measured physical property (*e.g.*, δ in NMR);
Y_0 the physical property of host solution before the guest is added;
Y_H the physical property (*e.g.*, δ_H in NMR) of the pure host **H**;
Y_G the physical property (*e.g.*, δ_G in NMR) of the pure guest **G**;
Y_{HG} the physical property (*e.g.*, δ_{HG} in NMR) of the pure **HG** complex;
Y_{HG_2} the physical property (*e.g.*, δ_{HG_2} in NMR) of the pure **HG$_2$** complex;
Y_{H_2G} the physical property (*e.g.*, δ_{H_2G} in NMR) of the pure **H$_2$G** complex;
ΔY $Y - Y_H$ (Y depends on mole fraction, *e.g.* NMR);
ΔY $Y - Y_0$ (Y depends on absolute concentration, *e.g.* UV-Vis);
$Y_{\Delta HG}$ $Y_{HG} - Y_H$;
$Y_{\Delta HG_2}$ $Y_{HG_2} - Y_H$;
$Y_{\Delta H_2G}$ $Y_{H_2G} - 2Y_H$;
δ observed NMR resonance;
δ_0 observed NMR resonance of host before the guest is added;
$\Delta\delta$ $\delta - \delta_0$;
δ_H NMR resonance of the free host **H**;
δ_{HG} NMR resonance of the complex **G**;
δ_{HG_2} NMR resonance of the complex **HG$_2$**;
δ_{H_2G} NMR resonance of the complex **H$_2$G**;
$\delta_{H_mG_n}$ NMR resonance of the complex **H$_m$G$_n$**;
$\delta_{\Delta HG}$ $\delta_{HG} - \delta_H$;
$\delta_{\Delta HG_2}$ $\delta_{HG_2} - \delta_H$;
$\delta_{\Delta H_2G}$ $\delta_{H_2G} - 2\delta_H$;
$\delta_{\Delta HG_i}$ $\delta_{HG} - \delta_H$ for resonance i ($i = 1, 2\ldots$);
ν_H the resonance frequency in hertz (Hz) of the free nucleus in the complex;
ν_{HG} the resonance frequencies in hertz (Hz) of the bound nucleus in the complex
A_{obs} the observed absorbance (UV-Vis);
A_{H_0} Absorbance of host solution before the guest is added;
ΔA_{obs} $A_{obs} - A_{H_0}$;
ε_H molar absorptivity of the free host **H**;
ε_{HG} molar absorptivity of the complex **HG**;
ε_{HG_2} molar absorptivity of the complex **HG$_2$**;
ε_{H_2G} molar absorptivity of the complex **H$_2$G**;
$\varepsilon_{\Delta HG}$ $\varepsilon_{HG} - \varepsilon_H$ (also called $\Delta\varepsilon$ in many texts);
$\varepsilon_{\Delta HG_2}$ $\varepsilon_{HG_2} - \varepsilon_H$;
$\varepsilon_{\Delta H_2G}$ $\varepsilon_{H_2G} - 2\varepsilon_H$;
F_{obs} the observed fluorescence;
F_0 fluorescence of host solution before the guest is added;
ΔF_{obs} $F_{obs} - F_0$;
I_0 intensity of excitation light;
Φ quantum yield;
b path-length in cm;
ε molar absorptivity;
k_X $I_0\Phi\varepsilon b$, proportionality constant for species **X**;
k_H^0 proportionality constant for the initial free host **H**;
k_H proportionality constant of the free host **H** (in the presence of a guest **G**);
k_{HG} proportionality constant of the complex **HG**;
k_{HG_2} proportionality constant of the complex **HG$_2$**;
k_{H_2G} proportionality constant of the complex **H$_2$G**
$k_{\Delta HG}$ $k_{HG} - k_H$
$k_{\Delta HG_2}$ $k_{HG_2} - k_H$
$k_{\Delta H_2G}$ $k_{H_2G} - 2k_H$
ΔH_{HG} molar enthalpy of formation for the 1 : 1 complex **HG**
ΔH_{HG_2} molar enthalpy of formation for the 1 : 2 complex **HG$_2$**;
ΔH_{H_2G} molar enthalpy of formation for the 2 : 1 complex **H$_2$G**;
α $K_1/4K_2$, interaction parameter;
p (Weber's) probability of binding;
y_{data} experimental fitted data points;
y_{calc} the calculated (fitted) data points;
y_{fit} $y_{data} - y_{calc}$, fitted residuals;
ss_{yy} $\Sigma(y_{data} - y_{calc})^2$, sum of squared fits;
SE_y standard error of the y-estimate;
χ^2 chi-squared value;
N number of data points;
k number of parameters to be fitted;
cov_{fit} the covariance of the fit;
$cov(y_{fit})$ the (co)variance of y_{fit};
$cov(y_{data})$ the (co)variance of y_{data};

Acknowledgements

The author is grateful to the Australian Research Council (ARC) for support by an Australian Research Fellowship/Discovery Project (DP0666325) and an additional ARC Discovery Project (DP0985059). The author also thanks the University of New South Wales (UNSW) for start-up funding and an UNSW GOLDSTAR grant for this research.

Notes and references

1 J.-M. Lehn, *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 89–112.
2 D. J. Cram, *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 1009–1020.
3 C. J. Pedersen, *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 1021–1027.
4 C. J. Pedersen, *J. Am. Chem. Soc.*, 1967, **89**, 2495–2496.
5 R. M. Izatt, J. H. Rytting, D. P. Nelson, B. L. Haymore and J. J. Christensen, *Science*, 1969, **164**, 443–444.
6 A. B. C. Deutman, C. Monnereau, M. Moalin, R. G. E. Coumans, N. Veling, M. Coenen, J. M. M. Smits, R. de Gelder, J. A. A. W. Elemans, G. Ercolani, R. J. M. Nolte and A. E. Rowan, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10471–10476.
7 P. Thordarson, R. G. E. Coumans, J. A. A. W. Elemans, P. J. Thomassen, J. Visser, A. E. Rowan and R. J. M. Nolte, *Angew. Chem., Int. Ed.*, 2004, **43**, 4755–4759.
8 K. A. Connors, *Binding Constants*, Wiley & Sons, New York, 1987.
9 J. E. A. Webb, M. J. Crossley, P. Turner and P. Thordarson, *J. Am. Chem. Soc.*, 2007, **129**, 7155–7162.
10 K. A. Connors, in *Comprehensive Supramolecular Chemistry*, ed. J. L. Atwood, J. E. D. Davies, D. D. MacNicol and F. Vögtle, Pergamon, Oxford, 1996, vol. 3, ch. 6, pp. 205–241.

11 H. Tsukube, H. Furuta, A. Odani, Y. Takeda, Y. Kudo, Y. Inoue, Y. Liu, H. Sakamoto and K. Kimura, in *Comprehensive Supramolecular Chemistry*, ed. J. L. Atwood, J. E. D. Davies, D. D. MacNicol and F. Vögtle, Pergamon, Oxford, 1996, vol. 8, ch. 10, pp. 425–482.
12 J. A. Goodrich and J. F. Kugel, *Binding and Kinetics for Molecular Biologists*, Cold Spring Harbour Laboratory Press, New York, 2007.
13 H. A. Benesi and J. H. Hildebrand, *J. Am. Chem. Soc.*, 1949, **71**, 2703–2307.
14 H. Lineweaver and D. Burk, *J. Am. Chem. Soc.*, 1934, **56**, 658–666.
15 R. L. Scott, *Recl. Trav. Chim. Pays-Bas*, 1956, **75**, 787–789.
16 C. S. Hanes, *Biochemistry*, 1932, **26**, 1406; J. B. Haldane, *Nature*, 1957, **179**, 832.
17 G. Scatchard, *Ann. N. Y. Acad. Sci.*, 1949, **51**, 660–672.
18 H. J. Motulsky and L. A. Ransnas, *FEBS J.*, 1987, **1**, 365–374.
19 H. J. Motulsky and 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.
20 K. A. Connors, A. Paulson and D. Toledo-Velasquez, *J. Org. Chem.*, 1988, **53**, 2023–2026.
21 B. Perlmutter-Hayman, *Acc. Chem. Res.*, 1986, **19**, 90–96.
22 H. L. Anderson, *Inorg. Chem.*, 1994, **33**, 972–981.
23 F. W. Dahlquist, *Methods Enzymol.*, 1978, **48**, 270–299.
24 C. A. Hunter and H. L. Anderson, *Angew. Chem., Int. Ed.*, 2009, **48**, 7488–7499.
25 G. Ercolani, *J. Am. Chem. Soc.*, 2003, **125**, 16097–16103.
26 K. A. Connors and J. A. Mollica, Jr., *J. Pharm. Sci.*, 1966, **55**, 772–780.
27 K. Kavallieratos, S.R. de Gala, D. J. Austin and R. H. Crabtree, *J. Am. Chem. Soc.*, 1997, **119**, 2325–2326.
28 H.-J. Schneider and A. Yatsimirsky, *Principles and Methods in Supramolecular Chemistry*, John Wiley & Sons, Chichester, 2000, p. 149.
29 C. S. Wilcox, in *Frontiers in Supramolecular Organic Chemistry and Photochemistry*, ed. H.-J. Schneider and H. Dürr, VCH, Weinheim, 1991, pp. 123–144.
30 G. Weber, in *Molecular Biophysics*, ed. B. Pullman and M. Weissbluth, Academic Press, New York, 1965, pp. 369–397.
31 P. Thordarson, E. J. A. Bijsterveld, J. A. A. W. Elemans, P. Kasák, R. J. M. Nolte and A. E. Rowan, *J. Am. Chem. Soc.*, 2003, **125**, 1186–1187.
32 J. Feeney, J. G. Batchelor, J. P. Albrand and G. C. K. Roberts, *J. Magn. Reson.*, 1979, **33**, 519–529.
33 C. S. Wilcox, J. C. Adrian, Jr., T. H. Webb and F. J. Zawacki, *J. Am. Chem. Soc.*, 1992, **114**, 10189–10197.
34 C. A. Parker and W. T. Rees, *Analyst*, 1962, **87**, 83–111.
35 J. M. Beechem, *Methods Enzymol.*, 1992, **210**, 37–54.
36 E. J. Billo, *Excel for Chemists*, Wiley-VCH, New York, 1997.
37 J. F. Rusling and T. F. Kumosinki, *Nonlinear Computer Modeling of Chemical and Biochemical Data*, Academic Press, San Diego, 1996.
38 D. B. Hibbert and J. J. Gooding, *Data Analysis for Chemistry*, Oxford University Press, New York, 2006.