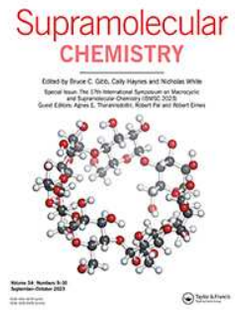


Supramolecular Chemistry

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gsch20

Driving to K-town: the quest for quality binding constants

Amar H. Flood, Douglas A. Vander Griend & Pall Thordarson

To cite this article: Amar H. Flood, Douglas A. Vander Griend & Pall Thordarson (2023) Driving to K-town: the quest for quality binding constants, Supramolecular Chemistry, 34:9-10, 320-325, DOI: 10.1080/10610278.2024.2359949

To link to this article: https://doi.org/10.1080/10610278.2024.2359949

SUPRAMOLECULAR CHEMISTRY
2023, VOL. 34, NOS. 9–10, 320–325
https://doi.org/10.1080/10610278.2024.2359949

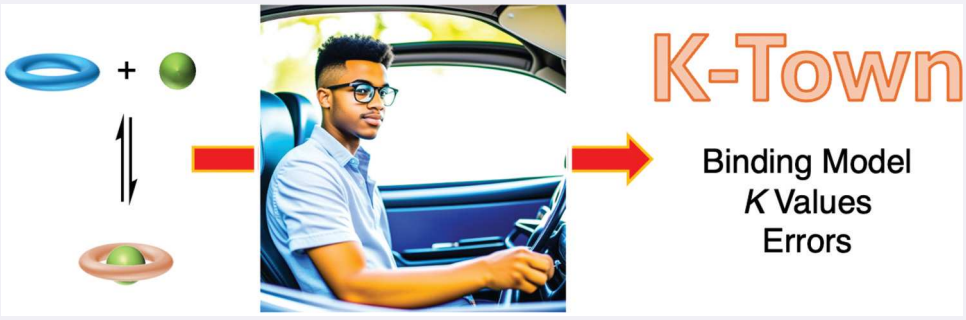
Driving to K-town: the quest for quality binding constants

Amar H. Flood [a], Douglas A. Vander Griend [b] and Pall Thordarson [c]

[a]Department of Chemistry, Indiana University, Bloomington, IN, USA; [b]Department of Chemistry and Biochemistry, Calvin University, Grand Rapids, MI, USA; [c]UNSW RNA Institute and the School of Chemistry, UNSW Sydney, Sydney, NSW, Australia

ABSTRACT

Binding constants (*K*) are foundational to supramolecular chemistry and quantified by modelling spectroscopic (NMR, UV-vis) titration data according to chemical equilibria. Spurred by growth in data science, the tools and methods for determining *K* values have accelerated in recent years. To share these advances, we provided a *Workshop on Quantifying Binding Constants* at ISMSC 2023 in Iceland and herein share the objectives, processes, and recommendations. We framed this short course in terms of learning to drive, from the basics 'under the hood', to 'behind the wheel', and navigating 'the open road'. These steps are crucial in the 'drive to K-town', where participants appreciate the importance of building, analysing, and comparing models. K-town is where they assess the hazards of incomplete models, inaccurate *K* values, and incorrect uncertainty assessment. We conclude with the Supramolecular Chemist's Pledge as a starting point for considering quality control in determining *K* values.



ARTICLE HISTORY
Received 2 May 2024
Accepted 20 May 2024

KEYWORDS
Binding constants; global analysis; data modelling; thermodynamic models; spectroscopic titrations

Introduction

Binding constants (K_a) are foundational to the field of supramolecular chemistry [1]. Knowledge of them helps us understand the forces that direct molecular associations and predict the outcomes of different reaction conditions. The quality of our insights rests squarely on the accuracy of our binding constants. Their accuracy relates to the quality of the raw titration data [2], and to the methodologies used to analyse, model and fit the data to binding constants [3]. As chemistry remains largely an experimental science, skills in data collection are nurtured and allowed to grow over many years of student training. By contrast, the skills needed for data analysis are not often taught. The growth of data science would suggest that the landscape is changing. This transformation is starting to have a positive impact on the determination of binding constants with the growth in the software tools, methods, and methodologies available for analysing titration data [4,5]. Many of these tools [6–9] are now findable, accessible, interoperable, and reusable thus complying with F.A.I.R. principles [10] but scattered through the literature. The motivation for running the *Workshop on Quantifying Binding Constants* was to collect these tools and best practices together, and to provide a means to share it with the community.

One key outcome of the recent developments in model building is the identification of sources of error [11]. From our experience as students and professionals, nothing beats repeat titrations, ideally in triplicate [4], for establishing errors and the limits of uncertainty in the binding constants we might determine. This source of error may be valid as the primary source in those cases where the binding event is a simple 1:1 association between host (H) and guest (G):

$$H + G \rightleftharpoons H \cdot G \quad K_a \qquad (1)$$

CONTACT Amar H. Flood aflood@indiana.edu Department of Chemistry, Indiana University, 800 E. Kirkwood Ave, Bloomington, IN 47405, USA

However, when the molecular association involves a higher order species being formed, e.g. 2:1 association complex, then the method of analysis also has an impact on the uncertainty. In such cases, equilibrium and stability constants can be determined for the stepwise and overall formation reactions. The broad goals of this workshop were to provide the philosophy, concepts, and tools needed to identify all reasonable binding models, evaluate the associated binding constants, and quantify the error from them (with strategies to minimise them). Thus, offering a pipeline (Figure 1) for the accurate determination of binding models and binding constants.

The specific goals of the workshop were to afford the participants with the ability to:

(1) Understand the quantitative aspects of chemical equilibria in supramolecular systems.
(2) Build and comprehensively analyse spectroscopic titration data.
(3) Use appropriate methodological tools to model data.
(4) Test different binding models.
(5) Design optimal experiments for quantifying binding constants.
(6) Report results in communally beneficial ways.

As a result, our intent was to educate and share best practices. We want those best practices to include a reflection on the limits of the analyses conducted. Specifically, to promote the idea that no single model of binding will perfectly recapitulate all the data. Thus, we recommend a best practice in which all analyses include a rationale for why one model was selected over others. The best way to provide the rationale is to compare the fitting for various models (Figure 1c). The best way to compare various models is from the mathematical quality of fit. These evaluations can then be used in combination with independent measures of the binding model, and the chemical sensibility of the findings.

Workshop summary and objectives

The first Workshop involved two 2-hour sessions of hands-on and lecture-style presentations (Figure 2). Participants included 17 students, postdocs and faculty members from nine countries (Germany, Australia, U.S.A., Belgium, Switzerland, Hong Kong, the Netherlands, Estonia and Iceland). Each participant brought a laptop for the hands-on portion. The Workshop was spread across two days with a homework portion completed overnight for review together on the second day.

Approach for the workshop

To achieve high quality binding constants takes understanding, skill, and the right computational tools. The workshop was designed to cover all the key experimental and methodological insights gathered over the years by the three instructors: Pall Thordarson, Douglas Vander Griend, and Amar Flood. The structure of the workshop consisted of three main parts that take their inspiration from learning to drive a car (Figure 3).

Part 1 looks 'under the hood' and focuses on the operational mathematics that defines equilibria and relates the spectroscopic signals to the chemical models.

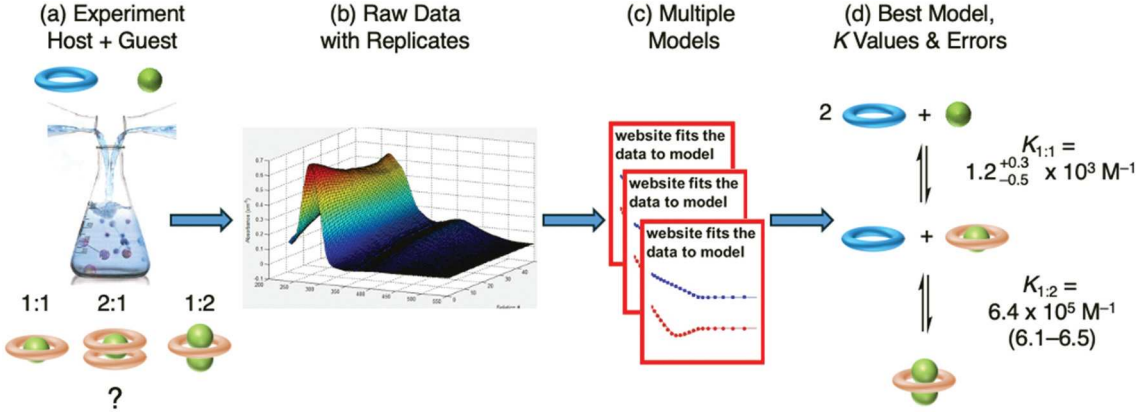


Figure 1. Pipeline for determination of the best binding models, binding constants and errors that provide the most accurate representation of the species and equilibria present when mixing hosts with guests. (a) Addition of host to guest could populate a range of species. (b) Replicate titrations and (c) tests of different combinations of reactions help (d) to identify best model, accurate *K* values and error estimates, which are modelled as asymmetric uncertainties for $K_{1:1}$ and a min-max range for $K_{1:2}$.

Figure 2. Photograph showing the hands-on and interactive aspect of the Workshop on Quantifying Binding Constants (DVG shown).

Figure 3. Learning good practices for determining binding constants at the Workshop is likened to learning to drive the car. Images created with www.kittl.com.

The *K* values dictate the extent of binding and consequently the concentration of all the species present. The Beer-Lambert Law quantifies the signal expected from each species present based on its concentration. All data can be harnessed, but only as long as the signals are additive. These fundamental relationships are familiar to chemists even if the matrix algebra required for the rigorous determination of *K* values from all the data (Figure 4) [12] may not be.

Without a basic understanding of the relationships between these quantities, analysis becomes a dangerous black box calculator. When modelling, one

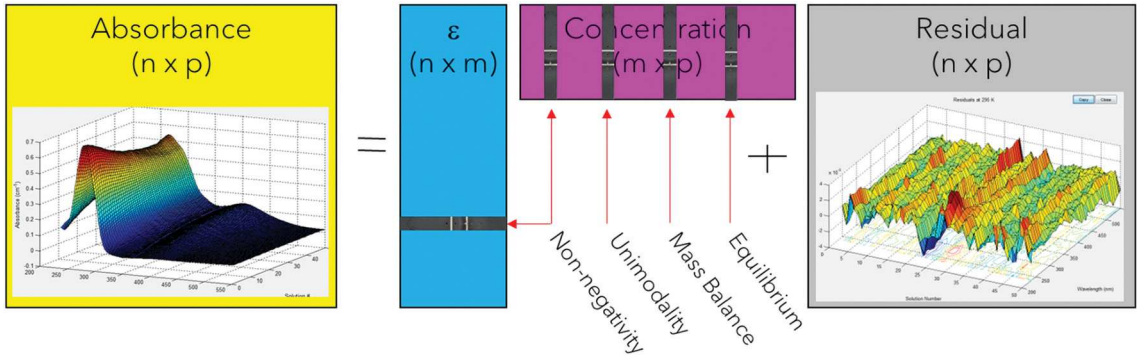


Figure 4. Beer-Lambert law in matrix form. Given a matrix of spectroscopic data, yellow (n is the number of data channels and p is the number of chemical solutions), find the non-negative molar absorption coefficients, blue (m is the number of distinct chemical species) and equilibrium concentrations (pink) that result in the smallest possible values in the residuals, gray, as defined by the root-mean-square residual (RMSR).

always gets an answer, so it is crucial to be able to evaluate that answer with insight. The tools enable people to avoid blind modelling. Understanding the mathematical connections also paves the way to re-designing titration experiments to optimise their sensitivity to the binding constants. Using our analogy, one can drive a car without understanding the inner workings of an internal combustion engine, but it will not be long before adding the right type of liquid to the fuel tank is critical to keep the car on the road.

We define *chemical model* here as the list of new chemical species that form from the association of the host and guest molecules. For example, the HG/HG_2 model, or the HG/H_2G model (Figure 1a). The *modeling* of data subsequently involves the optimisation of the corresponding binding constants for the chemical reactions that generate each species in the model. For example,

$H + G \rightleftharpoons HG \qquad K_{1:1}$

$HG + G \rightleftharpoons HG_2 \qquad K_{1:2}$

The modelling process also generates the corresponding spectral signatures for each species in the chemical model, whether molar absorptivity (UV) or the peak shift (NMR).

Part 2, aka 'behind the wheel', is focused on modelling titration data with the various computational tools available: Microsoft Excel, Supramolecular.org [6], Sivvu.org [7], and HypNMR [8]. Other programmes are also available. Each programme we addressed in the Workshop fits spectroscopic titration data according to a user-specified chemical model of equilibrating species (H, G, HG, HG_2, etc.) and quantifies the corresponding binding constants ($K_{1:1}$, $K_{2:1}$ etc.) that provide the best fit to the data. A final figure of merit, like root-mean-square-residual (RMSR), quantifies how well the model fits the data.

Spreadsheets like Excel are the most transparent in their calculations because the user can inspect every cell, but they are limited to simple models and the optimiser is clunky. Supramolecular.org is a website that was designed for NMR data but also works for UV-vis data and the like. It is also limited to simpler models, i.e. $H_n G_m$, n, $m \leq 2$, but is highly customised for chemists. Sivvu.org was designed for UV-vis data but also works for NMR data and can handle any model up to H_nG_m, n, $m \leq 6$. HypNMR can handle additional components in the model beyond the analyte and titrant, such as, off-target competition from ion pairing in addition to the targeted host-guest chemistry, e.g. H_nG_m, n, $m > 6$ as well as $H_nG_mP_pQ_q \ldots$, n, m, p, q … etc > 6. No matter the programme of choice, every single bit of the data can be modelled to provide a global fit. Therein, multiple wavelengths or peak shifts are used in the fitting because some are more sensitive to the different species than others. This global fitting allows for much greater resolution between possible models. The chemist no longer needs to assume simple one-to-one (1:1) binding or limit the model to a single chemical species. It is this variety of models (Figure 1c) that necessitates the third portion of the workshop.

Part 3, aka 'on the open road', deals with finding the best model out of many possible models (Figure 1d). The basic rule is to choose the model with the lowest RMSR that still has sensible spectroscopic signatures for each of the chemical species. Adding another species to the model will lower the RMSR, but if it does not correspond to a real species from the chemical soup, it will not exhibit a chemically sensible spectroscopic signature. For example, if the equilibrium concentration of an extraneous species is driven nearly to zero while its corresponding spectroscopic signature driven exorbitantly high, the species can be used to account for random noise in the data without detracting from the other parts of the fit.

Crystallographers follow a similar general approach, but the computational tools and the original quality of the data afford a lot less user input than fitting titration data. As a result, and akin to crystallography, the model can also be altered to match knowledge of the system acquired from independent means, e.g. diffusion NMR, NOE experiments etc. In such cases, the RMSR might be higher than a fit that is not constrained by the observations from complementary experiments.

Distinguishing between models

It can be particularly tricky to distinguish an HG/HG_2 model from an H_2G/HG model. If both chemical models fit the data well, we recommend designing a second (or third) titration experiment that best differentiates between these two realities. Perhaps, the starting concentration should be less or more. Perhaps, the range of equivalents titrated should be more. For any given chemical model, there is an optimal starting host concentration and an optimal range of guest equivalents to add to best support the equilibrium constants that govern the model. Check out the *Models* page of Sivvu.org for help in designing targeted titration experiments.

Empowering good science

The final part of the Drive to K-town involves the publication of reliable binding constants. Besides urging authors to make data available, which is facilitated by providing DOIs when using supramolecular.org and Sivvu.org, we also recommend three additional requirements:

(1) Report all the models tested, and at least, the two that are most competitive with each other. It is never sufficient to report a good standalone model. Every model yields an answer, so a model should only be reported as it is compared to other models. Reviewers and readers have a right to know which other models were not as good.
(2) Conduct replicate experiments. Even with reliable uncertainty ranges on each individual experiment, spectroscopic titration experiments tend to incorporate some measure of unknown or unmanageable physical/chemical distortion. Thus, replicate experiments in which the stock solutions are remade should be the expected standard since stock solution error is especially insidious. Add to this the fact that all sufficiently large binding constants fit the data alike[3] and there are plenty of systematic errors that can contaminate a dataset[5]. Thus, a replicate experiment might be better conducted at a different concentration regime (one or two orders of magnitude higher or lower) to simultaneously allow for distinguishing between competing models (see related section above).
(3) Assess the asymmetric uncertainty ranges on all published binding constants. Binding constants are non-linear parameters, e.g. titration curves are not linear, and the confidence intervals supported by any given dataset are not symmetric around the optimal *K* values. To determine the asymmetry in the uncertainty, researchers can quantify the minimum and maximum values of *K* that yield a 1% increase in RMSR. These ranges can also be simply calculated by bootstrapping, available through Sivvu.org, which quantifies the interval for which we can be 95% confident that the dataset supports the value of the binding constant. We suggest reporting uncertainty ranges as 1800^{+400}_{-600} M^{-1} or 1800 (1200 – 2200) M^{-1} or 1800 (+400, –600) M^{-1}, whether in the main body of a manuscript or in a footnote or in the supporting information.

These recommendations are captured in the Supramolecular Chemist's Pledge, which serves as a starting point for considering quality control in determining *K* values.

The "Supramolecular Chemist's" pledge

I will provide the most faithful binding model to describe the collection of equilibria and species present in solution across the conditions examined. I will determine the most accurate binding constants under the conditions used to quantify the association complexes. I will conduct and report the findings from at least one duplicate titration experiment, a test of at least one alternative binding model, and an estimate of the asymmetry in the model's uncertainty. I will hone my craft through communication with other practicing supramolecular chemists.

Looking to the future

Binding constants are key to characterising supramolecular systems, and the experiments we use to quantify them are, while straightforward, not trivial. Together we can improve our data-taking, data-modelling, and parameter-reporting using the tools available today. In the future, we look forward to more workshops that help more supramolecular chemists gainfully make the drive to K-town. Whether you need to brush up on what's 'under the hood', become comfortable 'behind the wheel', or accelerate 'on the open road', you should find something useful at each workshop.

The field of supramolecular chemistry has arrived at a place where we can quantify our binding constants with excellence and reliability to match the wonder and ingenuity of the chemical systems studied. Let's make the drive to K-town a regular road trip for the ages.

Acknowledgments

AHF acknowledges support from the Chemical Sciences, Geosciences, and Biosciences Division of the Basic Energy Sciences Program of the U.S. Department of Energy Office of Science (DE-SC0002728). DAV thanks the NSF for support (CHE 2004005).

Disclosure statement

DAV is cofounder of the non-profit Sivvu.org website. PT is the founder of the non-profit Opendatafit project that operates the supramolecular.org website.

Funding

The work was supported by the National Science Foundation [2004005]; U.S. Department of Energy [DE-SC0002728].

ORCID

Amar H. Flood http://orcid.org/0000-0002-2764-9155
Douglas A. Vander Griend http://orcid.org/0000-0002-8828-1112
Pall Thordarson http://orcid.org/0000-0002-1200-8814

References

[1] Thordarson P. Determining association constants from titration experiments in supramolecular chemistry. Chem Soc Rev. **2011**;40(3):1305–1323. doi: 10.1039/C0CS00062K

[2] Thordarson P. Binding constants and their measurement. In: Gale PA, Steed JW, editors. Supramolecular Chemistry. John Wiley & Sons, Ltd; 2012. p. 239–274.

[3] Kazmierczak NP, Chew JA, Michmerhuizen AR, et al. Sensitivity limits for determining 1:1 binding constants from spectrophotometric titrations via global analysis. J Chemomet. **2019**;33(5):e3119. doi: 10.1002/cem.3119

[4] Hibbert DB, Thordarson P. The death of the job plot, transparency, open science and online tools, uncertainty estimation methods and other developments in Supramolecular Chemistry Data Analysis. Chem Commun. **2016**;52 (87):12792–12805. doi: 10.1039/C6CC03888C

[5] Frassineti C, Ghelli S, Gans P, et al. Nuclear magnetic resonance as a tool for determining protonation constants of natural polyprotic bases in solution. Anal Biochem. **1995**;231(2):374–382. doi: 10.1006/abio.1995.9984

[6] Available from: http://app.supramolecular.org/bindfit/

[7] Available from: https://sivvu.org/

[8] Available from: http://www.hyperquad.co.uk/hypnmr.htm

[9] Available from: http://www.hyperquad.co.uk/hyss.htm

[10] Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. Comment: The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. **2016**;3 (1). doi: 10.1038/sdata.2016.18

[11] Kazmierczak NP, Chew JA, Vander Griend DA. Bootstrap methods for quantifying the uncertainty of binding constants in the hard modeling of spectrophotometric titration data. Anal Chim Acta. **2022**;1227:339834. doi: 10.1016/j.aca.2022.339834

[12] Wallace RM. Analysis of absorption spectra of multicomponent systems 1. J Phys Chem. **1960**;64 (7):899–901. doi: 10.1021/j100836a019