Summary about binding model and matlab model + data files in relation to:

Oligomeric phosphate clusters in macrocyclic channels

By

Subhamay Pramanik, Pall Thordarson, Victor W. Daya and Kristin Bowman-James

NMR titration data fitting to 1:3 binding models:
Input data is in a matlab file called:

kb4

The concentrations of the hosts are in vector = "*htot*", the guest one in "*Ltot*", both are combined in matrix "*initial*" and the raw observed NMR data is in "*DA*" (4 difference resonances were monitored, *i.e.*, this is a global fit approach).

The data was fitted to 1:3 using a similar approach to that described by Cozzolino and co-workers in their NMR study on a 3:1 complexation of a bis-antimony receptor with halide anions[1] and in a UV-Vis titration study on a nickel complex by Kudisch and co-works.[2] Here, the system of equations has been solved using the Wolfram Mathematica as shown previously in Kudisch *et al.*[2]

Full details on the terminology used here (H = host, G = guest, $[X]_0$ total concentration of species X, and K = association constants) for the binding models have been published previously.[2,3] Below the most important equations referred to in this paper are summarized:

For 1:3 H:G_3 complexation occurs according to equations S1-S3. Here, the stepwise association constants (K_1, K_2 and K_3) can be expressed in terms of the free energy changes (ΔG_1, ΔG_2 and ΔG_3) according to equations S4-S6 after correcting for statistical factors. If the binding was non-cooperative, ΔG_1, ΔG_2 and ΔG_3 would be expected to be the same, hence any difference between these are an indication of cooperativity.

$$K_1 = \frac{[\mathrm{HG}]}{[\mathrm{H}][\mathrm{G}]} \quad Eq.\ (S1) \qquad K_2 = \frac{[\mathrm{HG_2}]}{[\mathrm{HG}][\mathrm{G}]} \quad Eq.\ (S2) \qquad K_3 = \frac{[\mathrm{HG_3}]}{[\mathrm{HG_2}][\mathrm{G}]} \quad Eq.\ (S3)$$

$$\Delta G_1 = -R\mathrm{T}\ln\left(\frac{K_1}{3}\right) \qquad Eq.\ (S4)$$

$$\Delta G_2 = -R\mathrm{T}\ln(K_2) \qquad Eq.\ (S5)$$

$$\Delta G_3 = -R\mathrm{T}\ln(3K_3) \qquad Eq.\ (S6)$$

Equations S1 to S3, together with the corresponding mass-balance equations form the system of equations that is then solved using the Wolfram Mathematica program as mentioned above.[2] This yields a quartic (fourth-order) equation S7 for the concentration of the free guest [G]:

$$[G]^4(K_1K_2K_3) + [G]^3\{(K_1K_2)\} - (K_1K_2K_3[G]_0) + (3K_1K_2K_3[H]_0) + [G]^2\{(K_1)\} - (K_1K_2[G]_0) + (2K_1K_2[H]_0) + [G]\{1 - (K_1[G]_0) + (K_1[H]_0)\} - [G]_0 = 0 \quad Eq.\ (S7)$$

Binding models:

In the analysis below, four different binding models or "flavors"[4] are compared.

The first one is the stepwise (non-degenerate) "**full 1:3**" binding model. This model assumes three non-identical binding sites per molecule of host **1** that allows for cooperativity (negative or positive). Using a similar approach to the previously described 1:2 full model for NMR data,[3] we first define $\Delta\delta_{HG}$, $\Delta\delta_{HG_2}$ and $\Delta\delta_{HG_3}$ as the difference between the NMR chemical shifts ($\Delta\delta$) between the 1:1, 1:2 and 1:3 host-guest complexes, respectively and the molar absorptivity of the pure host (δ_H), e.g. for $\Delta\delta_{HG_3} = \delta_{HG_3} - \delta_H$. This, together with Equations S1-S3 and the mass balance equations yields S8 (with $\Delta\delta$ = change in NMR chemical shift upon performing the titration).

$$\Delta\delta = \left(\frac{\Delta\delta_{HG}K_1[G] + \Delta\delta_{HG_2}K_1K_2[G]^2 + \Delta\delta_{HG_3}K_1K_2K_3[G]^3}{1 + K_1[G] + K_1K_2[G]^2 + K_1K_2K_3[G]^3}\right) \qquad \text{Eq. (S8)}$$

Here the guest [G] concentration is obtained from the quartic equation S7 above.

We note also here that in practice, this is done in the matlab program in two steps; in the first step the concentrations of the 1:1, 1:2 and 1:3 complex are calculated and then a matrix division is performed to solve the system of linear equations that yields the molar absorptivtivies in S8.

*The second model (flavor) considered is the stepwise (non-degenerate) "**additive 1:3**" binding model.[4] To reduce the number of fitted parameters we note that in many circumstances it can be assumed that the NMR spectra of the 1:1, 1:2 and 1:3 are simply additive, i.e., for (each of) the chemical shifts observed, the change in chemical shift on going on from the free host to the 1:1 complex is simply 1/2 of the change between the 1:2 complex and the free host and 1/3 of the change between the 1:3 complex and the free host. It then follows that $\Delta\delta_{HG} = 2\,\Delta\delta_{HG_2} = 3\,\Delta\delta_{HG_3}$ and we can simplify equation S8 to yield equation S9.*

$$\Delta\delta = \left(\frac{\Delta\delta_{HG}\{K_1[G] + 2K_1K_2[G]^2 + 3K_1K_2K_3[G]^3\}}{1 + K_1[G] + K_1K_2[G]^2 + K_1K_2K_3[G]^3}\right) \qquad \text{Eq. (S9)}$$

*The third model (flavor) considered is the stepwise "**non-cooperative 1:3**" model.[4] Here we revert back to noting that the chemical shifts may not be correlated (see equation S7). However, instead we now make the assumption that the 1:3 complexation is non-cooperative[3] and, hence, after considering statistical factors, that $K_1 = 3K_2 = 9K_3$. If we define $K_{1n} = K_1 = 3K_2 = 9K_3$, from equation S7 we can simplify to obtain equation S10:*

$$\Delta\delta = \left(\frac{\Delta\delta_{HG}K_{1n}[G] + \Delta\delta_{HG_2}K_{1n}^2K_1K_2[G]^2 + \Delta\delta_{HG_3}K_{1n}^3[G]^3}{1 + K_{1n}[G] + K_{1n}^2[G]^2 + K_{1n}^3[G]^3}\right) \qquad \text{Eq. (S10)}$$

Here, K_{1n} replaces the need to fit 3 parameters (one for each step-wise binding constants) and once the data has been fitted to S10, the three step-wise non-cooperative binding constants can be readily obtained $K_{1n} = K_1 = 3K_2 = 9K_3$.

*The fourth and last model (flavor) is the "**statistical 1:3**" model.[4] Here we not only make the assumption, that the binding is non-cooperative ($K_1 = 3K_2 = 9K_3$) but also that the chemical shifts are all additive ($\Delta\delta_{HG} = 2\Delta\delta_{HG_2} = 3\Delta\delta_{HG_3}$). We can now combine the approached used for S9 and S10 to get equation S11:*

$$\Delta\delta = \left(\frac{\Delta\delta_{HG}(K_{1n}[G] + 2K_{1n}^2K_1K_2[G]^2 + 3K_{1n}^3[G]^3)}{1 + K_{1n}[G] + K_{1n}^2[G]^2 + K_{1n}^3[G]^3}\right) \qquad \text{Eq. (S11)}$$

Compared to the "full" model which has 3 binding constants and 3 independent chemical shifts per proton resonance observed (which could be many for a global fit) we have now a much simpler model with only one binding constant K_{1n} and one molar absorptivity change $\varepsilon_{\Delta HG}$ (per wavelength point) to fit to our data.

How we fit the models:

Using equations S8-S11 we wrote a code in the Matlab program similar to that described in our earlier work.[2-4] The program files are all uploaded on the github repository:

https://github.com/pallithordarson/Oligophosphate-1-3-model/

This includes a file that solves the quartic equation S7 – this file is named:

nmr1to3bbbkb.m

And then a pair of files with the prefixes (file extension is the matlab .m extension):

Runfitbinding1to3nmr[SUFFIX] and *nmr1to3fitbindk[SUFFIX]*

With the former starting the program, the latter is then called upon from that program for the iteration process. The [SUFFIX] of these file names then varies depending on which of the four models / equations above they refer to:

- [SUFFIX] = kb "full model" Equation S8
- [SUFFIX] = aggkb "additive model" Equation S9
- [SUFFIX] = noncoopkb "non-cooperative model" Equation S10
- [SUFFIX] = statkb "statistical model" Equation S11

After fitting the data to these models, a matlab data file (.mat) with all the inputs was generated.

This package includes a total of 4 files; 4 for each of the 4 models. The names of these file are created using this system:

Prefix = *fitted1to3*

Sub-suffix = which model, i.e. *full*, *agg*, *noncoop* and *stat*

Ending = *x4* (denotes this is a global fit from 4 different chemical resonances).

Where full = full model, agg = additive model, noncoop = non-cooperative and stat = statistical model.

As an example, the file name:

fitted1to3noncoopx4.mat

corresponds to the data fitted to the non-cooperative binding model (equation S10).

Comparing the models:

To analyze and compare the model used we used two indicators for the quality of the fit; the covariance of the fit (cov_{fit}) as in our previous work,[2-4] and the Bayesian Information Criteria (BIC)[5] as another robust method for model comparison. The challenge is that the more complicated the model (= higher number of parameters), the better the fit is likely to be. However, generally we should only pick a more complicated model if the fit is significantly better when compared to a simpler model. As discussed previously, the cov_{fit}

is relatively insensitive to the number of parameters used and if there is a more than 2-3-fold reduction in cov_{fit} then the more complicated model (flavor) is a reasonable one. The BIC is a different type of measure of fit and is based on the calculated log-likelihood, the number of parameters, and the number of data points, whereby an increase in the number of parameters leads to a penalty (increase) in the BIC value. A low BIC is generally better and when comparing two models, and a difference of more than 6-10 is usually considered as a strong evidence that there is a significant difference between the two models being compared.

In this study, the "full" 1:3 model (flavor) always appeared to fit the experimental data considerably better than any of the other models (flavors). This comparison is shown in detail in an excel files that have been included as:

BowanResultsfor1to3modelall4

Overall results for the full 1:3 binding model:

(repeated from the SI with the manuscript for clarity – also shown in the excel file above)

Table S2. Chemical shift for amide (NH), methylene (b and a) and methyl (c) fitted with the four proton resonances using the "full" (unconstrained) 1:3 model. After correcting for statistical factors this corresponds to binding energies of ΔG_1 = -6.2 kJ mol^{-1}, ΔG_2 = -12.0 kJ mol^{-1} and ΔG_3 = -8.6 kJ mol^{-1} indicating some positive cooperativity (binding gets stronger after 1:1 complex is formed).

	Statistical analysis				binding constants		
Model (flavour)	BIC (Bayesian)[a]	ΔBIC compared to full model[b]	Data Points	Degrees of freedom	K_1 / M^{-1}	K_2 / M^{-1}	K_3 / M^{-1}
full	-559.03	n/a	80	65	**968**	**68204**	**1024**
additive	-429.56	-129.47	80	73	912	1084	3 x 10^{-8}
Non-cooperative	-513.03	-46.00	80	67	12551	4184[c]	1395[c]
statistical	-380.78	-178.26	80	75	12143	4048[c]	1349[c]

[a]Bayesian Information Criteria (BIC).[5] The lower BIC, the better. BIC can be negative and if so, the more negative the better.

[b]Defined here as BIC(X model)-BIC(full model). A ΔBIC > 6-10 is usually considered as strong evidence in favor of the model with a lower (more negative) BIC.

[c]By default, in these models $K_1 = 3K_2 = 9K_3$.

Reference

1. Qiu, J.; Song, B.; Li, X.; Cozzolino, A. F., Solution and gas phase evidence of anion binding through the secondary bonding interactions of a bidentate bis-antimony(iii) anion receptor. *Physical Chemistry Chemical Physics* **2018,** *20*, 46-50

2. Kudisch, J.; Lim, C.-H.; Thordarson, P.; Miyake, G. M. M., Energy Transfer to Ni-Amine Complexes in Dual Catalytic, Light-Driven C–N Cross-Coupling Reactions, *Journal of the American Chemical Society*, **2019**, *141*, 19479-19486.

3. Thordarson, P., Determining association constants from titration experiments in supramolecular chemistry. *Chemical Society Reviews* **2011,** *40* (3), 1305-1323.

4. Howe, E. N. W.; Bhadbhade, M.; Thordarson, P., Cooperativity and Complexity in the Binding of Anions and Cations to a Tetratopic Ion-Pair Host. *Journal of the American Chemical Society* **2014,** *136*, 7505-7516.

5. Schwarz, G., Estimating the Dimension of a Model. *The Annals of Statistics* **1978,** *6* (2), 461-464