ELSEVIER

The determination of equilibrium constants for simple reaction systems from temperature-dependent physicochemical data

R.J. Willson[a,*], A.E. Beezer[b], M.A.A. O'Neill[b]

[a] *GlaxoSmithKline, St. George's Avenue, Weybridge, Surrey KT13 0DE, UK*
[b] *School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX, UK*

Abstract

This paper provides a pragmatic approach for calculation of equilibrium constants that have been determined from systems at equilibrium by measuring a component of the equilibrium at several (minimum of three) different temperatures. The paper builds on previous work that has shown by selecting appropriate temperatures so that the van't Hoff isochore could be written in terms of the total amount of a measured analyte (A_{total}) associated with the equilibrium, equilibrium constants could be determined. This paper describes a mathematical method that uses the linearity of the van't Hoff plot, $\ln K$ versus $1/T$ as the basis to systematically test for an appropriate A_{total} value and derive the best values of equilibrium constants. It will also show that this method is freed from constraints in choosing the study temperatures.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Equilibrium constants; MathCad; van't Hoff; Gravimetric vapour sorption; Solubility

1. Introduction

Studies that are concerned with equilibrium are usually expected to conclude by reporting the equilibrium constant. This reaction parameter can give access to a great deal of information about a process studied. For example, $\Delta_R G$ and $\Delta_R H$ can be determined directly from van't Hoff type approaches [1]. Knowing $\Delta_R G$ and $\Delta_R H$, $\Delta_R S$ can be found. A previous publication [2] included a brief discussion of the application of the van't Hoff equation to the analysis of calorimetric data. A general approach was proposed for the calculation of equilibrium constants, K, from calorimetric data which in turn was used to obtain $\Delta_R G$ and $\Delta_R H$.

This paper extends the approach and is concerned with the development of a strategy for the determination of equilibrium constants from limited temperature-dependent experimental data inputs (not necessarily calorimetric) for simple reaction systems. Indeed, the approach has already been used

in a study of water vapour uptake onto a drug surface [3]. Briefly, the form of the equations used in the development of the strategy is illustrated in the reaction scheme shown below. Later, it will be shown that this strategy can be applied to more complicated systems.

2. Development of the equations

For a system at equilibrium such that it follows the relationship:

$$A \leftrightarrow A–B$$
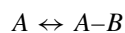
where A interacts with an excess of substrate B to form a complex $A–B$, an associated equilibrium constant can be expressed by:

$$K = \frac{A_{eq}}{A–B_{eq}} \tag{1}$$

(Strictly activities should be employed here but, in the Henry's Law region, it is assumed that activity coefficients

---

* Corresponding author. Tel.: +44 1932 822116.
  E-mail address: Richard.J.Willson@GSK.com (R.J. Willson).

are equal to 1, hence, quantities are adequate to describe the equilibrium.) Eq. (1) can be simplified if the quantity terms are written for a common component of the equilibrium. For example, the total amount of A reacting in the system is:

$$A_{\text{total}} = A_{\text{eq}} + A - B_{\text{eq}} \tag{2}$$

where A_{total} is the total quantity of a component A in the system and A_{eq} is the quantity at equilibrium with A–B_{eq}.

Eq. (1) can, therefore, be written as:

$$K = \frac{A_{\text{eq}}}{A_{\text{total}} - A_{\text{eq}}} \tag{3}$$

The analytical measurement of A_{eq} in Eq. (3) can, of course, be made by using any analytical method where the measured response is linearly proportional to the quantity of A_{eq}. For example, absorbance where quantity is proportional to absorbance through $\text{Abs} = \varepsilon Al$ (where there is a linear relationship in the Beer–Lambert Law).

As described in a previous publication [3], A_{total} can be found if the study is performed at a minimum of three different temperatures. In earlier treatments, it was shown that, providing the selected temperatures (T_1, T_2 and T_3, with corresponding measurement A_1, A_2 and A_3), satisfy the conditions described in the following section and that the van't Hoff isochore relates equilibrium constants and temperature through (for the three temperatures considered here):

$$\ln \frac{K_1}{K_2} = \frac{-\Delta H}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)$$

and

$$\ln \frac{K_2}{K_3} = \frac{-\Delta H}{R}\left(\frac{1}{T_2} - \frac{1}{T_3}\right)$$

then provided that the right hand sides of each of the above equations are equal then it follows that

$$\ln \frac{K_1}{K_2} = \frac{K_2}{K_3}. \tag{4}$$

The equality is achieved by recognising that ΔH and R are constants and that the temperatures are related as:

$$\frac{T_2 - T_1}{T_1 T_2} = \frac{T_3 - T_2}{T_2 T_3} \tag{5}$$

for $T_1 = 293$ K and $T_2 = 298$ K, $T_3 = 303.2$ K and T_4 must be equal to 308.5 K, etc.

Hence, A_{total} can be made the subject of the equation:

$$A_{\text{total}} = \frac{A_2^2 A_1 + A_2^2 A_3 - 2A_3 A_2 A_1}{A_2^2 - A_3 A_1} \tag{6}$$

where A_1, A_2 and A_3 are the measured quantity of A at the temperature T_1, T_2 and T_3. This means that the value of A_{total} does not have to be known, it can be calculated from the above equation.

These mathematical treatments detailed in previous publications [2,3] are shown here to emphasis the dependence

A_{total} has on the quality of measured A_{eq} values. Inspection of Eq. (6) shows that values of A_{eq} are made up of cubic multiples of A values. Even though the input experimental data may be excellent (accurate, precise), error magnification in this calculation represents a real hazard. It is necessary, therefore, to manage the data so that "best possible" values for Ks are determined. This can be done through a MathCad® [4] worksheet as outlined later.

Having determined A_{total}, the equilibrium constants at each temperature can be determined. From the values of K, other parameters can be determined: $\Delta_R G = -RT \ln K$, $\Delta_R H$ is determined from the slope of the $\ln K$ versus $1/T$ plot (slope $= -\Delta_R H/R$; intercept $= \Delta_R S/R$).

The development outlined above requires that, over the temperature range studied, the system conforms to the van't Hoff isochore relationship, i.e. $\ln K$ is linear with $1/T$.

This relationship may fail to hold if: (i) there is a mechanism change within the temperature range of interest and/or (ii) $\Delta_R H$ is not independent of temperature.

However, it is clear that the test for appropriateness is based upon conformity of the experimental data with the van't Hoff relationship.

3. Data management

A MathCad® algorithm was written to optimise the calculated A_{total} value by iteration of A_{total} over a defined range of values. The best value of A_{total} is chosen on the basis that the resulting K-value (see Eq. (3)), determined at each temperature, conform as closely as possible to van't Hoff's theorem.

The principle of the analysis is to measure A_{eq} at a minimum of three different temperatures. A_{eq} values are then used with a series of ascribed A_{total} values to construct a vector consisting of a string of K-values (calculated from Eq. (3)).

In the algorithm, A_{total} is a range variable starting at A_{max} (i.e. the maximum A_{eq} measured at any of the study temperatures) and ending at an arbitrary number that is considerably larger than A_{max} (see below for clarification).

A matrix is thus constructed of K-values that is nT columns wide and mA_{total} rows long (where n represents the number of temperatures used and m is the number of A_{total} values tested). The mathematical form of the matrix is given in Eq. (7).

The range of A_{total} chosen should start at A_{max} (i.e. A_{total} should have a number that is equal to or greater than any of the measured A_{eq} values in the study) and increase to the maximum theoretical absorbency, i.e. the total amount of A that is physically added in the system. For practical application, a wide range is chosen with a small number of intervals to gain an approximate value of A_{total}. Having gained an approximate A_{total} value, the range can then be reduced and the interval number increased to converge at the best A_{total} value. This method has an added advantage in that A_{total} is selected from a string of values that are derived independently of temperature. Therefore, it is not a requirement that A_{eq}

values are measured at predefined temperatures, as in the case where A_{total} is calculated using Eq. (6).

$$\text{Matrix} = \begin{vmatrix} \text{for } p \in 1 \ldots \text{Row} \\ \quad \text{for } z \in 1 \ldots \text{Col} \\ \qquad K_{p,z} \leftarrow \frac{A_z}{\mathrm{d}t_p - A_z} \\ K \end{vmatrix} \qquad (7)$$

Here, A_z is the measured A_{eq} at temperatures of "z". $\mathrm{d}t_p$ is a range variable and consists of "p" values for A_{total}. In this example, a matrix containing values of K that is "Col" columns wide and "Row" rows deep would be constructed.

Each row in the matrix is made up of K-values derived at each of the different temperatures and with each of the A_{total} value. Testing for linearity is achieved by calculation of the linear regression coefficient for $\ln K$ with the associated $1/T$ data for each row of K-values. The row that has a linear regression coefficient closest to 1 indicates the most appropriate value of A_{total}.

4. Method robustness

Three different types of datasets were simulated to test the robustness of the algorithm: a dataset that was made up of two different processes: a dataset where $\ln K$ was not linearly proportional to $1/T$ and a dataset with the addition of a random error to simulated measuring error. The aim of these simulations was to explore the types of datasets that might be expected from "real" systems and to test the ability of the mathematical program to recover the simulated equilibrium constants. To construct the simulated data, K-values were determined from the van't Hoff equation using ascribed values for $\Delta_R H$ and $\Delta_R S$ and a series of T values. As a reminder, a linear plot of $\ln K$ versus $1/T$ has the slope of $-\Delta_R H/R$ and an intercept of $\Delta_R S/R$ (where R is the gas constant). Having determined K-values from $\Delta_R H/R$ and $\Delta_R S/R$, ascribing a value to A_{total} allows a series of simulated measured A_{eq} values. A_{eq} was then used as absorbance values (ε was given a value of 1 in the Beer–Lambert Law). The simulated datasets consisted of eight data points, absorbance versus temperature.

4.1. Datasets simulating a change in mechanism

Datasets made up of two processes were simulated by joining two separate, linear datasets each containing four values of $\ln K$ and $1/T$. It is assumed that a change in mechanism would not affect A_{total} (see comments about A_{total} later). Ascribing a value for A_{total}, the simulated equilibrium constants were then used in Eq. (3) to determine the corresponding A_{eq} and then absorbance and together with temperature they were entered into the Mathcad program. Various combinations of data were constructed; where the simulated absorbance ranged from 0.1 to 1 and from 0.98 to 0.99, simulating extremes of sensitivity; where the linearity (R^2) of the com-
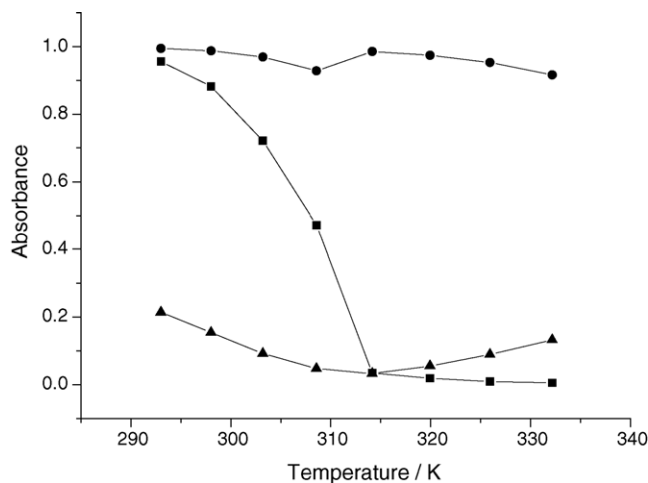

Fig. 1. Simulated absorbance vs. time datasets where there is a change in mechanism during the observation period. (●) represents the case where both mechanisms have a small change in absorbance, (■) where one mechanism has a large change in absorbance and (▲) where the mechanism starts off exothermic and switches to endothermic.

bined dataset (of $\ln K$ versus $1/T$) ranged from 0.98 to 0.44, exploring how the degree of change in mechanism affected the results and for datasets that start off exothermic and end endothermic. Selections of the extremes of these simulated data are shown in Fig. 1. The program was then run to determine if the initial values of K could be returned. In all cases, the equilibrium constant could be returned without any significant error (less than 1%), see Fig. 2.

4.2. Datasets were $\ln K$ versus $1/T$ is not linear

These data were simulated using Eq. (8), that is $\ln K$ is expressed as an exponential function of T.
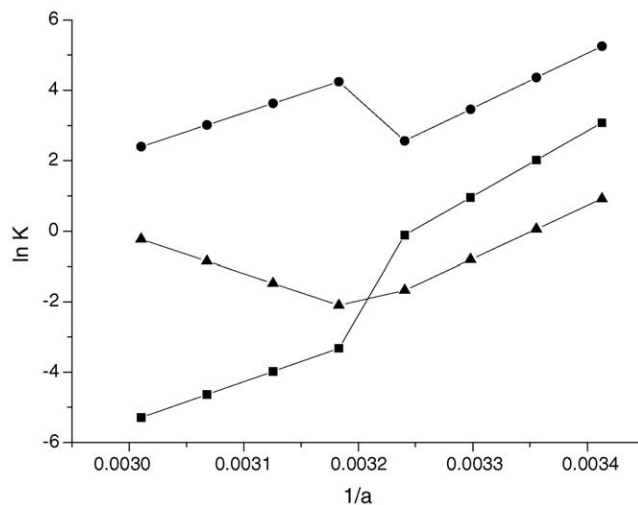
$$\ln K = A \exp(-B \times T) \qquad (8)$$

Fig. 2. The resulting calculated equilibrium constants derived from the absorbance vs. temperature datasets shown in Fig. 1 are plotted here in their linear form, $\ln K$ vs. $1/T$. The symbols used correspond to Fig. 1.
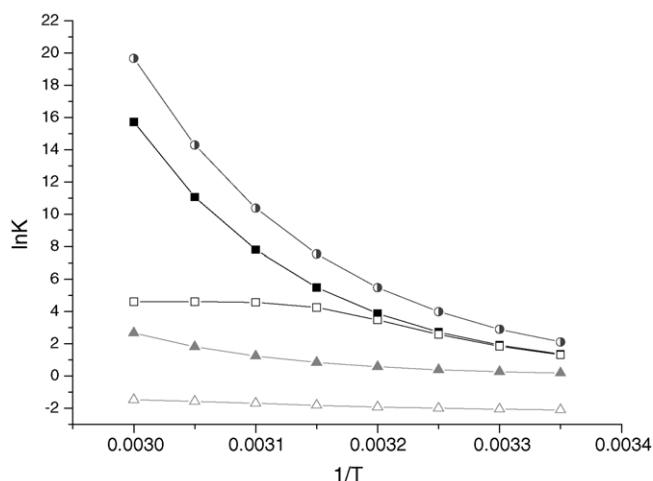
Fig. 3. shows the plot of $\ln K$ vs. $1/T$ for a series of datasets with varying degrees of curvature as determined from the linear regression coefficient, R. Open symbols represent the calculated equilibrium constants. The linear regression coefficients for the simulated datasets (closed symbols) are: ($\bullet$) 0.954, ($\blacksquare$) 0.947 and ($\blacktriangle$) 0.937.

A and B are constants arranged to produce typical absorbance values in conjunction with Eq. (3) with an assigned value of A_{total}. Various datasets were constructed with different curvature, determined from a linear regression coefficient (R). It was found that the program could return the correct K-values providing R was no worse than 0.954, see Fig. 3.

4.3. Datasets with a random error added to simulate measuring error

Absorbance versus T datasets were constructed from values of ΔH, ΔS and A_{total}. A percentage error, ranging from 1 to 10%, was selected. The error was applied to a dataset as a range $\pm$ of the chosen percentage error, randomized between 0 and the number corresponding to the percentage error chosen. Thus, an error of 10% applied to an absorbance value of 1 gave a random number between 0.9 and 1.1.

K data was constructed for eight temperatures using the following values:

$\Delta_{\text{R}} H = 68 \, \text{kJ mol}^{-1}$, $\Delta_{\text{R}} S = 229 \, \text{J mol}^{-1}$ and a temperature range of 293.15–332.28 K. Converting to absorbance

Table 1
Selection of equilibrium models that can be incorporated into the Mathcad program

$A \rightleftharpoons B$	$K = A_{\text{eq}}/A_{\text{total}} - x$
$2A \rightleftharpoons B$	$K = A_{\text{eq}}/(A_{\text{total}} - A_{\text{eq}})^2$
$A + B \rightleftharpoons Z$	$K = A_{\text{eq}}/(A_{\text{total}} - A_{\text{eq}})(B_{\text{total}} - A_{\text{eq}})$
$2A \rightleftharpoons Y + Z$	$K = (Y - A_{\text{eq}})(Z - A_{\text{eq}})/(A - A_{\text{eq}})^2$

using the reaction model shown in line 1 of Table 1, an addition of $\pm 5\%$ random error was added. Six different datasets were constructed and analysed, the results of the analysis are shown in Table 2.

5. Practical exploitation of the strategy

5.1. Application of the MathCad analysis to gravimetric vapour sorption

In a recent publication [3] showing the application of GVS to calculation of drug–vapour equilibrium constants, mass changes were measured at various water vapour partial pressures at three different temperatures. From these data, an equilibrium constant was determined and it was shown that a plot of $\ln K$ versus $1/T$ is linear over the temperature range studied. These data were subsequently analysed using the mathematical approach above. Unsurprisingly, this study also results in a linear relationship for a plot of $\ln K$ versus $1/T$, however, the important conclusion is that the slope of the line for the van't Hoff type analysis in the GVS study is comparable to the "optimised" slope resulting from this study. Table 3 summarises the results of this study.

A second comparison was made of this method for calculation of equilibrium constants for the equilibrium between dimeric rhomboid and trimeric hexagon supra molecules [5]. In this study, it was found that an equilibrium existed between dimeric and trimeric macromolecules. The equilibrium was determined by measuring the ratio of the dimer and trimer by ^{1}H NMR at temperatures 253, 273, 293, 313, 333 and 353 K. A correction was made for the activity coefficient in this study; however, the observed equilibrium constants were also reported. The published data showing speciation as a function of temperature was analysed using the procedure in this paper

Table 2
Six simulated datasets were constructed from the K-values shown in the second column of the table above

Simulated A_{eq} values with 5% random error added	Simulated K-values	MCAD results, mean K ($N = 6$)	S.D.	T (K)
0.44697	0.41244	0.41455	0.01091	283.15
0.62038	0.68083	0.68733	0.02008	288.15
0.77989	1.12389	1.10507	0.0199	293.33
0.97804	1.85525	1.8154	0.08702	298.70
1.15582	3.06253	3.11821	0.16513	304.27
1.25721	5.05544	5.19391	0.18575	310.05
1.3231	8.34529	8.33495	1.01663	316.05
1.38801	13.77592	12.43221	2.28913	322.28

From these values, corresponding absorbance values were determined to which a 5% random error was added. The calculated equilibrium constants, column 3, were calculated from the absorbance vs. temperature data.

Table 3
A comparison of published equilibrium vapour pressures [3] and the analysis of the mass change data in the Mathcad program

0.2 P/P_0 literature values	0.2 P/P_0 MCAD	0.3 P/P_0 literature values	0.3 P/P_0 MCAD	0.4 P/P_0 literature values	0.4 P/P_0 MCAD	0.6 P/P_0 literature values	0.6 P/P_0 MCAD
0.271	0.270	0.244	0.246	0.232	0.241	0.220	0.221
0.127	0.127	0.123	0.124	0.121	0.125	0.117	0.118
0.060	0.059	0.062	0.063	0.063	0.065	0.062	0.063

Table 4
A summary of published data [5] showing calculated equilibrium constants for dimeric and trimeric macromolecules at equilibrium

T (K)	% Trimer	K_{obs}	K determined by MCAD
253	72.5	2.63636	2.3833
273	67.5	2.07692	1.905703
293	61	1.5641	1.455153
313	57	1.32558	1.241289
333	52	1.08333	1.02121
353	48.7	0.94932	0.898193

The % speciation data was put through the Mathcad program and the resulting calculated equilibrium constants are reported.

and the results are given in Table 4. Fixing trimer$_{max}$ at 100% resulted in K-values that, unsurprisingly, match the published values of K_{obs}. If MathCad is allowed to optimise the value of trimer$_{max}$, the program returned an optimised trimer$_{max}$ values of 103%. The resulting van't Hoff plot of the optimised data gave a linear regression coefficient of 0.99883 compared to the regression coefficient of 0.99881 for the raw data.

6. Discussion.

The application of the algorithm requires that the mathematical form of the equilibrium can be written, and hence, is known. The model described in Eq. (7) is for a simple equilibrium where $A \Leftrightarrow A$–B. Table 1 gives other, but not exhaustive, forms of mathematical expression for reaction models that can be entered into Eq. (7). For example data taken from Cox and Hayman [6] that describe the equilibrium of ClO dimers indicates an equilibrium process of the type:

$$2ClO \rightleftharpoons (Cl_2O_2)$$
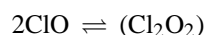
The mathematical form of this process is shown in Eq. (8), below. The data provided by Cox and Hayman [6] was also tested and for the nine temperatures published, the Mathcad results were consistent with the literature values.

$$K = \frac{A_{eq}}{(A_{total} - A_{eq})^2} \tag{8}$$

The meaning of A_{total} in Eq. (3) is not the total amount of A in the system but the total amount of A that is involved in the equilibrium. That is, A_{total} is not necessarily the same as the quantity of A added to the system. For a solid in equilibrium with its solution, in the case of solubility experiments, one

may be tempted to write:

$$K = \frac{[A]_{aq}}{A_{solid}} \tag{9}$$

In this context, A_{solid} has an activity coefficient of 1 and, therefore, K is equal to $[A]_{aq}$. This term is sometimes referred to as the solubility product [7].

However, A_{solid} can be written as:

$$A_{solid} = A_{total} - [A]_{aq} \tag{10}$$

Therefore,

$$K = \frac{[A]_{aq}}{A_{total}} - [A]_{aq}$$

Clearly K is not directly proportional to $[A]_{aq}$ as not all the solid A in the system is involved with the equilibrium.

The various simulations outlined in Section 4 are intended to show the robustness of the method to handle "real" data where there will be some degree of analytical error. The effects of outliers were simulated (data not shown here). Where a single data point was increased by 40%, no significant error in the returned equilibrium constants were noted, except for the equilibrium constant corresponding to the outlier. The error was reduced if the values of ln K are interpolated from a linear fit of calculated ln K versus $1/T$.

7. Conclusion

The MathCad® algorithm has been constructed to allow the calculation of equilibrium constants from a set of data where a quantity at equilibrium is determined as a function of temperature. Where the relationships between K and T for a system at equilibrium conform to the van't Hoff isochore, this method will select the most appropriate value of A_{total} to give the most linear relationship between ln K and $1/T$. The essence of the algorithm is to construct a range variable of A_{total} values. For each A_{total}, an equilibrium constant is calculated from the measured A_{eq} and the value of A_{total}. As A_{eq} is measured at a minimum of three temperatures, the resulting K is tested for linearity using a van't Hoff approach. This process continues for every value of A_{total} in the range. The most linear van't Hoff analysis gives the optimum associated K-values.

Where a change in mechanism occurs over the temperature range studied, providing A_{total} is not changed and that there is a linear relationship between ln K and $1/T$ throughout, then

the equilibrium constants can be returned. Where $\ln K$ is not a linear function of $1/T$, the equilibrium constants can be returned providing that the deviation from linearity does not result in a linear regression coefficient (R) of less than 0.954. This is, of course, dependent on the range of the X and Y experimental values selected.

Random error, when applied to the datasets unsurprisingly results in the poorest returned equilibrium constants were the change in absorbance over the observed temperature range is small. To maximize the accuracy of this method, a good separation of the measured parameters should be sort by using appropriate concentrations or temperature range. Datasets simulated from a combination of these potential sources of error were not tested. Accuracy can also be improved with greater precision in the analytical measurement. For the example given above, if the error added to the simulated A_{eq} data is changed from 5 to 1%, the error associated with the calculated K-values diminish considerably.

Reports in the literature that detail measured values of an equilibrium process, determined at several different temperatures, are somewhat scarce. However, the study used for this paper together with previous sorption analysis and simulated data has shown the method to be robust and reliable. As with most types of mathematically modeled studies, some knowledge of the mechanism is required so that a mathematical form of the model can be written into the mathematical algorithm.

References

[1] E.B. Smith, Basic Chemical Thermodynamics, Clarendon Press, Oxford, 1990, ISBN 0-19-855564-4.
[2] A.E. Beezer, A.C. Morris, M.A.A. O'Neill, R.J. Willson, A.K. Hills, J.C. Mitchell, J.A. Connor, J. Phys. Chem. B 105 (2001) 1212.
[3] R.J. Willson, A.E. Beezer, Int. J. Pharm. 258 (2003) 77.
[4] http://www.mathcad.com.
[5] T. Yamamoto, A. Arif, P. Stang, J. Am. Chem. Soc. 125 (2003) 12309.
[6] R.A. Cox, G.D. Hayman, Nature 332 (28) (1988) 796.
[7] A.T. Florenve, D. Attwood, Physicochemical Principles of Pharmacy, The MacMillan Press, London, 1982, ISBN 0-33-323405-7.