J. Iran. Chem. Soc., Vol. 5, No. 4, December 2008, pp. 522-534.

JOURNAL OF THE
Iranian
Chemical Society

The Investigation of Equilibria in Solution

M. Maeder[a] and H. Abdollahi[b,*]
[a]*Department of Chemistry, University of Newcastle, Australia*
[b]*Institute of Advanced Studies in Basic Sciences, Zanjan, Iran*

(Received 12 July 2008, Accepted 25 October 2008)

This contribution describes the practical and computational aspects relevant for the quantitative analysis of chemical equilibria. First a systematic nomenclature based on stability constants is introduced. Next, practical aspects of potentiometric pH titrations and spectrophotometric titrations are discussed. The description of the computational aspects includes the Newton-Raphson algorithm that allows the computation of all species concentrations for a given model, set of formation constants and total component concentrations. The second computational part introduces the Newton-Gauss algorithm for non-linear data fitting. Three practical examples illustrate all the above. They include the spectrophotometric investigations of the interaction of Bi(III) with Cl^-, of Cu(II) with EDTA and the potentiometric investigation of the Zn(II) ethylenediamine system. All Matlab softwares for the data generation and analysis are available from the authors.

Keywords: Chemical equilibrium, Titration, Non-linear least-squares fitting, Protonation equilibria, Complexation equilibria

INTRODUCTION

A large class of interactions between molecules and ions in solution can be classified as Lewis acid-base interactions. The quantitative investigation of these interactions encompasses two parts: kinetics and equilibria. The kinetic studies are focused on the determination of the mechanism of interaction and the rates for all individual reaction steps, while the equilibrium studies concentrate on the determination of the species that are formed at equilibrium together with the quantitative determination of the strength of the interactions as defined by equilibrium or formation constants.

In this tutorial paper we will concentrate on equilibrium studies. As most interactions are investigated in water, emphasis is put on aqueous solutions. We will start with an introduction into the theory of equilibria in solution, the law of mass action, the notations required for the quantitative description of the interactions, which can be rather complex, particularly in aqueous solutions [1,2]. This will be followed by a discussion of the types of measurements that can be carried out and the nature of data that are delivered [3]. Next, we will introduce the computational methods are required for the analysis of the measurements. In order to give the reader the possibility to apply all the above concepts, we will use three practical examples: the spectrophotometric investigation of the interactions of Bi^{3+} with Cl^-, of the interaction of *edta* with Cu^{2+} ion, and the potentiometric determination of the interaction of Zn^{2+} ion with ethylenediamine, *en*. A suite of Matlab programs that perform the analysis of the above data sets is available from the authors.

CHEMICAL EQUILIBRIUM IN SOLUTION

Most examples of chemical equilibria can be seen as

*Corresponding author. E-mail: abd@iasbs.ac.ir

interactions between Lewis acids and Lewis bases. In most instances the Lewis acid is a metal ion (*M*) and we use the corresponding word ligand (*L*) for the Lewis bases. However, it is important to remember that there are many other examples of Lewis acids and bases, *e.g.* the proton is a Lewis acid and water molecules of hydration are Lewis bases. The chemical equation describing the 1:1 interaction between metal ion and ligand.

$$M + L \xrightleftharpoons{K_{ML}} ML \tag{1}$$

and the law of mass action states that

$$K_{ML} = \frac{[ML]}{[M][L]} \tag{2}$$

where the charges are omitted for simplicity.

In aqueous solution, the protons are always present and more importantly they are also Lewis acids which can compete with any other Lewis acid present. The protonation equilibrium can be described in an analogous way as:

$$H + L \xrightleftharpoons{K_{HL}} HL \tag{3}$$

with

$$K_{HL} = \frac{[HL]}{[H][L]} \tag{4}$$

All metal ions can interact with all ligands (Lewis bases) present in the solution and many ligands have several Lewis base sites (they are polydentate) and therefore there is often a long list of different species that can be formed between metals, protons and ligands. For instance, there are complexes with 6 ligands, *e.g.*, $Co(NH_3)_6^{3+}$or ligands that are protonated several times, *e.g.*, $edtaH_6^{2+}$ (Note, we omit charges for general species such as ML but include them for specific ones such as $edtaH_6^{2+}$).

It is obviously important to develop a nomenclature that uniquely describes all the species formed in solution with their appropriate equilibrium constants. Let us concentrate on an example of a metal-ligand equilibrium study in aqueous solution. Using the nomenclature commonly employed in coordination chemistry, there are three components, *M*, *L* and *H* present. As an example, the metal can be Cu^{2+} to interact with the ligand *ethylenediamine*. In aqueous solution, these components interact to form the following species, *HL*, H_2L, *ML*, ML_2, ML_3, *MLH*, MLH_{-1} and *OH*. In fact, many more species may be formed, *e.g.* ML_2H_{-1}, but they only form in small and unobservable concentrations. Note the expressions used: the components are the basic units that interact with each other to form the species; it is convenient to include the components in the list of species. Each of the species is formed by the appropriate number of components, and the quantitative relationship between the component and species concentrations is defined by the formation constant. The general equation is;

$$\begin{aligned} &mM + lL + hH \xrightleftharpoons{\beta_{mlh}} M_mL_lH_h \\ &\beta_{mlh} = \frac{[M_mL_lH_h]}{[M]^m[L]^l[H]^h} \end{aligned} \tag{5}$$

The above example is represented in Table 1. The structure of most of the complexes listed in Table is clear, some structures, maybe not. Figure 1 gives a few examples.

The formation constant β_{11-1} represents the equilibrium for the formation of the deprotonated complex MLH_{-1}.

$$M + L - H \xrightleftharpoons{\beta_{11-1}} MLH_{-1} \tag{6}$$

Different, chemically more intuitive, ways of defining this species are given in Eq. (7); however, it is not easy to define the equilibrium constants for such equilibria in a consistent way, the standard notation is more transparent and generally applicable.

$$\begin{aligned} &ML \rightleftharpoons ML(OH) + H \\ &\text{or} \\ &ML - H \rightleftharpoons ML(OH) \end{aligned} \tag{7}$$

The definition of the stabilities of all species as a function of the component concentrations *via* the β_{mlh} values is consistent and allows the development of general, compact and thus fast computer programs for data analysis. Any stability constant, *e.g.* the first one shown in Eq. (7), can be expressed as a

Table 1. Notation for Equilibrium Modelling

Species	Notation			Formation constant
	m	*l*	*h*	
M	1	0	0	$\beta_{100} = 1$
L	0	1	0	$\beta_{010} = 1$
H	0	0	1	$\beta_{001} = 1$
LH	0	1	1	$\beta_{011} = \frac{[LH]}{[L][H]}$
LH_2	0	1	2	$\beta_{012} = \frac{[LH_2]}{[L][H]^2}$
ML	1	1	0	$\beta_{110} = \frac{[ML]}{[M][L]}$
ML_2	1	2	0	$\beta_{120} = \frac{[ML_2]}{[M][L]^2}$
ML_3	1	3	0	$\beta_{130} = \frac{[ML_3]}{[M][L]^3}$
MLH	1	1	1	$\beta_{111} = \frac{[MLH]}{[M][L][H]}$
MLH_{-1}	1	1	-1	$\beta_{11-1} = \frac{[MLH_{-1}]}{[M][L][H]^{-1}}$
H_{-1}	0	0	-1	$\beta_{00-1} = [OH][H] = K_W$

function of the standard formation constants:

$$K = \frac{[ML(OH)][H]}{[ML]} = \frac{[ML(OH)]}{[M][L][H]^{-1}} \frac{[M][L]}{[ML]} = \beta_{11-1}\ \beta_{110}{}^{-1} \qquad (8)$$

THE TITRATION EXPERIMENT

There are two goals in an equilibrium investigation: (i) the determination of the model, which is essentially the list of different species that are formed from the components, and (ii) the evaluation of the formation constants for all species. The simplest case is the determination of a formation constant for an equilibrium of the type represented in Eq. (1), where the model is known. One can prepare a solution with known total concentrations of *M* and of *L*, measure the concentrations one of the species ([*M*] or [*L*] or [*ML*]), compute the concentrations of other species *via* the law of conservation of mass and then compute K_{ML} by substituting concentrations of all species into Eq. (2). However, the model is usually not known and many more experiments are required in order to determine the model and evaluate the formation constants.

The experiment for the investigation of an equilibrium consists of the preparation of a series of solution with different known total concentrations of the components. Usually this is called a titration. Obviously there is no limit to the number of possible experiments. Apart from the very simplest cases, a titration will always include more data than minimally required for the determination of a particular constant. This increases the robustness of the analysis and also delivers statistical information about the results, such as standard deviations of the fitted parameters (see below). Usually this is done as a titration: a solution of one of the components is added stepwise to the solution of the other component(s). The added solution is delivered by a burette, which can be manual or automatic under computer control. After the addition of each aliquot and enough time for the establishment of the equilibrium, the mixture is analyzed in some appropriate way. The scalar *nmeas* represents the number of additions and

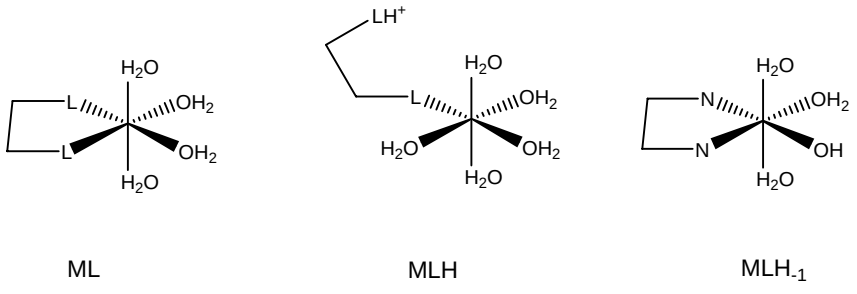

Fig. 1. Reasonable structures for the complexes *ML*, *MLH* and MLH_{-1}.

measurement that were taken during the titration.

Titrations can be undertaken manually, by preparing a series of solutions with known total concentrations; it is much more efficient and reliable to use computerized titration set-ups. Here a computer controls the burette and the pH-meter or spectrophotometer. After each addition, the solution is stirred and enough time is allowed for complete equilibration before the data are acquired.

In principle any measurement that provides information about at least one of the species can theoretically be employed for a titration. The most common titration uses the pH electrode which provides information on the free proton concentration, [*H*], at any time during the titration. Also of importance are the spectrophotometric titrations where the absorption spectrum of the solution is acquired as a function of the progress of the titration. Absorption data are governed by Beer-Lambert's law; they are only indirectly related to the concentrations *via* the usually unknown molar absorptivities for all absorbing species. The advantage of spectrophotometric titrations is the availability of absorption spectra for all species; this allows structural analysis of the species. No such information is available from pH titrations. There are other ion specific electrodes that allow potentiometric titrations, but none of these electrodes is anywhere near as useful as the pH electrode. Their range is very limited. Several other spectroscopies are potentially applicable; examples include NIR, IR, ESR and NMR titrations.

It is clear that experimental conditions, *i.e.* the range of total concentrations of the components, have to be chosen carefully. Any species for which the formation constant is to be determined needs to exist at some minimal concentration somewhere during the complete titration. If a protonation equilibrium occurs around pH 10, a titration that covers only the pH range from say pH 1 to pH 7 will not contain the information required to determine the above protonation constant. Unfortunately, the range of required experimental conditions is not always as obvious. This is naturally the case when novel systems are investigated as the results are not known prior to the design of the experiment. There are only a series of titrations, starting with general conditions and subsequent titrations, where the conditions can be chosen according to the results of the primary titrations.

In aqueous solution, protonation and complexation are usually coupled and generally a series of titrations is required to define all protonation and complexation constants. Usually, a minimum of 2 titrations are required. The first titration is a titration of a solution of the protonated ligand with the base, resulting in the protonation constants of the ligand; the second is titration of an acidic solution of the ligand and metal with base, delivering additional information about the complexation equilibria.

If ternary complexes are investigated, the components may be *M*, *L'*, *L''* and *H*, or *M'*, *M''*, *L* and *H* or even *M''*, *M''*, *L'*, *L''* and *H*. In such instances a more extensive series of titrations is required to define all formation constants.

The original experimental data required for the subsequent analysis include the total concentrations of the components in the solutions after each addition, conveniently collected as rows in a matrix $\mathbf{C}_{\mathbf{tot}}$, and the actual measurements which are collected in a vector $\mathbf{d}_{\mathbf{meas}}$ in the case of pH titrations, or as rows of a matrix $\mathbf{D}_{\mathbf{meas}}$ for spectrophotometric titrations.

DATA ANALYSIS

Computer programs written for the analysis of measurements for equilibrium investigations contain two parts that require specific attention.

The more obvious one is the algorithm for parameter fitting, its task is to determine the optimal values for the parameters for a given measurement and model. In a titration experiment the parameters to be fitted are usually the formation constants and, in the case of a spectrophotometric titration, additionally the molar absorption spectra of all absorbing species. Titrations are also used in analytical applications, and then the concentrations of some of the components can also be fitted to the titration data.

The other important part is the computation of all species concentrations for a given set of total component concentrations and formation constants. This calculation has to be performed for the solution after each addition of reagent during the complete titration. This second task forms the core of the data fitting and, therefore, we will discuss it first.

The Newton-Raphson Algorithm

The task of the Newton-Raphson algorithm is the computation of the species concentrations for a given set of

formation constants and total concentrations of the components, $[M]_{tot}$, $[L]_{tot}$ and $[H]_{tot}$. The basis for these computations is the law of mass action and the law of conservation of mass. At equilibrium, the different species of the model all reach a certain concentration and these concentrations have to obey the above laws, *i.e.* the law of mass action has to relate the species concentrations to the free concentrations of the components and, as an example, the total concentration of metal that exists in the different species has to equal the known total concentration $[M]_{tot}$.

These computations are usually fairly complex and iterative algorithms have to be employed [4]. While there are many possible ways to do this, the standard algorithm starts with initial guesses for the free component concentrations. Theses are used to compute all species concentrations and subsequently mass conservation is checked. If there are any discrepancies, the iterative algorithm is continued.

It is probably best to illustrate the algorithm based on the example given in Table 1. Each species concentration is computed from the formation constants and the free component concentrations as given in Eq. (9) for a general equilibrium and for the example of MLH

$$[M_{m_i}L_{l_i}H_{h_i}] = \beta_{m_i l_i h_i}[M]^{m_i}[L]^{l_i}[H]^{h_i} \qquad (9)$$

e.g. $[MLH] = \beta_{111}[M][L][H]$

Obviously Eqs. (9) must be met for all species. Thus, for *nspec* species formed there are *nspec*-3 such equations. (Recall, the 3 components themselves are also species)

The hydroxide ions and deprotonated species need some additional attention. The concentration for the species $[MLH_{-1}]$ is computed as:

$$[MLH_{-1}] = \beta_{11\text{-}1}[M][L][H]^{-1} \qquad (10)$$

The formation constant for the hydroxide ions is:

$$\beta_{00-1} = \frac{[OH^-]}{[H^+]^{-1}} = [OH^-][H^+] = K_w \qquad (11)$$

It is most convenient to define the hydroxide concentration as negative proton concentration, $[OH^-] = -[H^+]$. In aqueous solution, the addition of x moles of OH^- is equivalent to removing x moles of H^+.

For each of the components, M, L and H, we can write the following equations:

$$\begin{aligned}
&[M]_{tot\text{-}calc} = [M] + [ML] + [ML_2] + \\
&[ML_3] + [MLH] + [MLH_{-1}] \\
&[L]_{tot\text{-}calc} = [L] + [LH] + [LH_2] + \\
&[ML] + 2[ML_2] + 3[ML_3] + [MLH] + [MLH_{-1}] \\
&[M]_{tot\text{-}calc} = [H] + [LH] + 2[LH_2] + [MLH] - \\
&[MLH_{-1}] - [OH]
\end{aligned} \qquad (12)$$

where $[M]_{tot}$, $[L]_{tot}$ and $[H]_{tot}$ are the known independent variables in a titration. They are computed from the reagent solutions and dilutions occurring during the titration. See the chapter *The Titration Experiment*. The total concentrations of the components are stored in the matrix $\mathbf{C_{tot}}$. These total concentrations have to equal the sums over all appropriately weighted species concentrations, Eq. (12). The differences are collected in a vector **d**. The goal of the algorithm is to determine the free component concentrations such that the differences **d** are zero.

$$\begin{aligned}
d_M &= [M]_{tot} - [M]_{tot_calc} \\
d_L &= [L]_{tot} - [L]_{tot_calc} \\
d_H &= [H]_{tot} - [H]_{tot_calc}
\end{aligned} \qquad (12)$$

An important advantage of using the formation constant formalism is that equations such as Eq.

$$\begin{aligned}
[M]_{tot_calc} &= \sum_{i=1}^{nspec} m_i\beta_i[M]^{m_i}[L]^{l_i}[H]^{h_i} \\
[L]_{tot_calc} &= \sum_{i=1}^{nspec} l_i\beta_i[M]^{m_i}[L]^{l_i}[H]^{h_i} \\
[H]_{tot_calc} &= \sum_{i=1}^{nspec} h_i\beta_i[M]^{m_i}[L]^{l_i}[H]^{h_i}
\end{aligned} \qquad (13)$$

Note the formation constants for the components themselves are all = 1, *i.e.*, $\beta_{100} = \beta_{010} = \beta_{001} = 1$. This simplifies the

notation, the writing of computer code and also reduces computation time.

It is easy to see that we are dealing with a set of *nspec* equations, 3 Eqs. (13) and *nspec*-3 Eqs. (9), with *nspec* unknowns, [*M*], [*L*], [*H*], [*LH*], …, [MLH_{-1}]. While not all systems of *n* equations with *n* unknowns have a solution, it can be shown that the systems we are dealing with always have exactly one physically possible solution. It is of course the one that is realised the actual solution in the titration vessel.

As with most non-linear system of equations, there is no explicit solution. There are many ways of solving such a system of equations. The Newton-Raphson algorithm is usually well behaved and is relatively straightforward to implement. The algorithm is sketched in the flow sheet of Fig. 2 which to a large extent is self explanatory.

The Jacobian **J** is the *ncomp* × *ncomp* matrix of the derivatives of the differences **d** with respect to the component concentrations [*M*], [*L*], [*H*]. A shift vector that moves the vector of component concentrations towards the true solution is computed subsequently. The iterative process is usually well behaved and converges well, however, there is no guarantee for conversion and special measures need to be introduced to deal with such cases.

These calculations are performed for each step of the titration, each one resulting in a vector of species concentrations. It is most convenient to collect all these vectors as rows in a matrix **C** in which each column contains the concentration profile for one species. In summary one can state that the Newton-Raphson algorithm computes the species concentration matrix **C** from the matrix $\mathbf{C}_{tot}$ of total component concentrations, based on the equilibrium model used.

The explanations given here are insufficient for complete understanding and more extensive explanations are beyond the scope of this tutorial, for details refer to [4]. Study of the Matlab program supplied by the authors, see end of this paper, can also assist the comprehension of the methods discussed.

FITTING

The measured data, $\mathbf{D}_{meas}$ (or $\mathbf{d}_{meas}$), as distinguished from theoretically perfect data, are always corrupted by experimental errors, instrumental shortcomings, noise and *etc*.

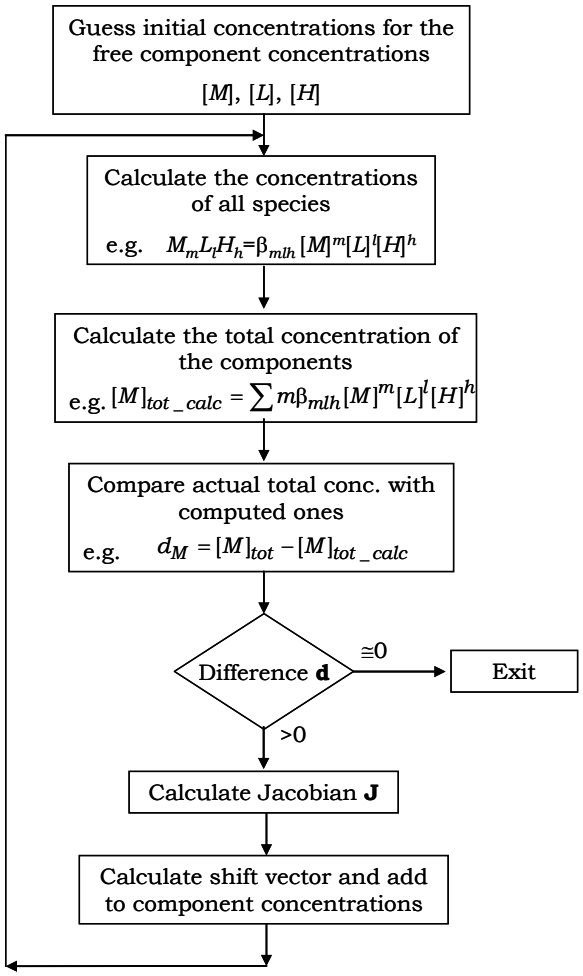

Fig. 2. Flow sheet for the Newton Raphson algorithm, see text for details

The true data are never known. The idea of data fitting is to determine a calculated set of data, $\mathbf{D}_{calc}$, which resembles the measured data as closely as possible. This calculated data set is defined by the model (both chemical and type of measurement) and the collection of parameters, the vector **par**. The differences between the measured and calculated data are the residuals, **R**:

$$\mathbf{R} = \mathbf{D}_{meas} - \mathbf{D}_{calc}(\text{model}, \mathbf{par}) \qquad (14)$$

Note that for univariate data, the residuals **r**, and the data $\mathbf{d}_{meas}$ and $\mathbf{d}_{calc}$ are vectors instead of matrices, otherwise there are no

differences. The task of the fitting algorithm is to find a, hopefully unique, set of parameters for which the measured data $\mathbf{D}_{\mathbf{meas}}$ and their calculated values $\mathbf{D}_{\mathbf{calc}}$ are as similar as possible. The measure for the quality of the fit is the sum of squares, *ssq*, which is the sum over the squares of all elements of the matrix **R** (or the vector **r**)

$$ssq = \sum_i \sum_j R_{i,j}^2 \tag{15}$$

Combining Eqs. (14) and (15) clearly demonstrates that *ssq* is a function of the parameters, and of course of the model and the data themselves.

The chemical model, together with the parameters allows the computation of all concentrations for the complete measurement, *i.e.* all species concentrations as a function of the titration. This is done by the Newton-Raphson algorithm, as described above. The results of these calculations are collected in the matrix **C**.

Concentrations are very rarely measured directly; measured data are usually only indirectly related to the concentrations. In this contribution we discuss in some details two types of titrations, namely, potentiometric pH titrations and spectrophotometric titrations. In pH titrations the measurement is a vector $\mathbf{d}_{\mathbf{meas}}$ which is a record of the pH of the solution. The calculated $\mathbf{d}_{\mathbf{calc}}$ is -log([H^+]), note that the proton concentration is collected in one particular column of the concentration matrix **C**. The data collected in spectrophotometric titrations are more complex and their analysis requires additional attention.

In spectrophotometric titrations, according to Beer-Lambert's law, the relationship between the concentrations and the measurements can be expressed as a matrix equation

$$\mathbf{D}_{\mathbf{meas}} = \mathbf{CA} + \mathbf{R} = \mathbf{D}_{\mathbf{calc}} + \mathbf{R} \tag{16}$$

$\mathbf{D}_{\mathbf{meas}}$ has the dimensions *nmeas* × *nlam*, where *nlam* is the number of wavelengths at which data were acquired. The molar absorptivities of all species at the measured wavelengths are collected in a matrix **A** of dimensions *nspec* × *nlam*. The dimensions of the matrix **C** are *nmeas* × *nspec*.

The parameters describing the chemical model are collected in a vector **par** in Eq. (14). The matrix **A** of molar absorptivities that relates the concentrations to the measurement are linear parameters. The non-linear parameters **par** require an iterative algorithm that starts with initial guesses and, hopefully, converges towards the optimal solution in a reasonable number of iterations and amount time. Linear parameters can be 'fitted' in explicit equations, so that no iterative process is required. Given $\mathbf{D}_{\mathbf{meas}}$ and **C**, the best estimate for **A** can be calculated explicitly as:

$$\mathbf{A}=\mathbf{C}^{+}\mathbf{D}_{\mathbf{meas}} \tag{17}$$

$\mathbf{C}^+$ is called the pseudo inverse of **C**. ($\mathbf{C}^+$ can be calculated as $\mathbf{C}^+ = (\mathbf{C}^t\mathbf{C})^{-1}\mathbf{C}^t$ but there are numerically better algorithms available [4-6].

Multivariate, or multi-wavelength data are modelled by both non-linear and linear parameters; usually by a few non-linear ones, the vector **par**, and many linear ones, the matrix **A** in Eq. (16). It is crucial to deal with them separately, iteratively refining only the non-linear parameters while dealing with the linear ones with explicitly. As demonstrated in Eq. (18), *ssq* can be defined as a function of the non-linear parameters **par** only; the linear parameters are effectively eliminated.

$$\begin{aligned} &\mathbf{C} = f(\text{model}, \mathbf{par}) \\ &\mathbf{A} = \mathbf{C}^{+}\mathbf{D}_{\mathbf{meas}} \\ &\mathbf{R} = \mathbf{D}_{\mathbf{meas}} - \mathbf{CA} = \mathbf{D}_{\mathbf{meas}} - \mathbf{CC}^{+}\mathbf{D}_{\mathbf{meas}} \\ &ssq = \sum_i \sum_j R_{i,j}^2 = f(\text{model}, \mathbf{par}) \end{aligned} \tag{18}$$

The non-linear parameters to be fitted are the equilibrium constants of the considered system that define the matrix of concentration profiles **C**. The equilibrium constants are refined so as to minimize the sum of squares of the residuals matrix, *ssq*.

The least-squares fitting of non-linear parameters is necessarily an iterative process. There are several algorithms available for that task. The most commonly used one is Newton-Gauss method [4,7]. There are algorithms that are conceptually simpler, but in very few instances are such alternatives more efficient.

The Newton-Gauss method is fast, relatively straightforward to implement and it delivers estimates for the

standard deviations of the fitted parameters. The equations governing the Newton-Gauss algorithm are based on a truncated Taylor series expansion of the sum-of-squares as a function of the parameters. The Newton-Gauss algorithm for *ssq* minimization requires the computation of the derivatives of the residuals with respect to the parameters. These derivatives are collected in the Jacobian **J**. Again, we refer to more specialised text for more extensive explanations [4,7].

$$\mathbf{J} = \frac{\delta \mathbf{R}}{\delta \mathbf{par}} \tag{19}$$

The computation of **J** is relatively time consuming but the reward is the speed of convergence which is quadratic close to the minimum. For equilibrium investigations it is usually possible to compute the Jacobian explicitly [4], however, it is always possible to approximate **J** numerically:

$$\frac{\delta \mathbf{R}}{\delta par_i} \cong \frac{\mathbf{R}(\mathbf{par} + \Delta par_i) - \mathbf{R}(\mathbf{par})}{\Delta par_i} \tag{20}$$

In this equation (**par** $+\Delta par_i$) is a new parameter vector with only the *i* -the parameter par_i shifted by the small amount Δpar_i. Typically Δpar_i is calculated as $1 \times 10^{-4}\ par_i$.

The iterative refinement of the parameters is given by the following formula. The shift vector $\boldsymbol{\Delta}$**par** is computed and added to the vector **par**.

$$\boldsymbol{\Delta}\mathbf{par} = -\mathbf{J}^{+}\mathbf{R} \tag{21}$$

Usually this results in convergence, and the minimum in *ssq* is reached in a few iterations. The test for convergence is performed by comparing the calculated *ssq* with the value from the previous iteration. If improvement is below a certain threshold, *i.e.* the shift in the parameters resulted in no further improvement of the *ssq* value, then the process is terminated and the results are reported. In the case of divergence (increase in *ssq* from one the next iteration) the well-proven Marquardt-Levenberg algorithm is invoked [4-6]. The pseudo-inverse $\mathbf{J}^{+}$ is calculated as $\mathbf{J}^{+} = (\mathbf{J}^{t}\mathbf{J})^{-1}\mathbf{J}^{t}$ and the Marquardt parameter *mp* is added to the diagonal elements of ($\mathbf{J}^{t}\mathbf{J}$) prior to inversion. Increasing the Marquardt parameter shortens the shift vector and directs it to the direction of steepest descent.

An additional useful property of the Newton-Gauss algorithm is that it allows direct estimation of the errors in the non-linear parameters [7]. The inverted Hessian matrix $\mathbf{H}^{-1} = (\mathbf{J}^{t}\mathbf{J})^{-1}$, without the Marquardt parameter added, is the variance-covariance matrix of the parameters. The diagonal elements contain information on the parameter variances and the off-diagonal elements the covariances. The standard error σ_{par} of the fitted parameters par_i can be estimated from the expression:

$$\sigma_{par_i} = \sigma_R \sqrt{d_{i,i}} \tag{22}$$

where $d_{i,i}$ is the *i*-th diagonal element of the inverted Hessian matrix $\mathbf{H}^{-1}$. $\acute{\sigma}_R$ represents the estimated standard deviation of the measurement error in $\mathbf{D}_{\mathbf{meas}}$ or $\mathbf{d}_{\mathbf{meas}}$.

$$\sigma_R = \sqrt{\frac{ssq}{df}} \tag{23}$$

where *df* is the degree of freedom, which is defined as the number of experimental values *m* (elements of **D** or **d**), subtracted by the number of optimised parameters *np*, $df = m - np$.

Figure 3 is a flow diagram showing the Newton-Gauss-Levenberg/Marquardt (NGL/M) method [4].

It is worth noting that the determination of the correct chemical model for a given measured process is a completely different and a much more difficult task than fitting the parameters of a given model. There are so-called model free analysis methods which can be applied to the spectrophotometric titrations which can give some preliminary insight into the complexity of the data. In some instances it is possible to get good estimates for the concentration profiles and the absorption spectra of the interacting species. As before, we refer to the original literature for more details.

One thing, however, is clear from the beginning: the smaller the number of parameters, the easier the task of fitting them. And this is why it is so important to separate the linear from the non-linear parameters, as described in Eq. (18).

APPLICATIONS

In this section a number of examples are given where the techniques described have been applied to three simulated data

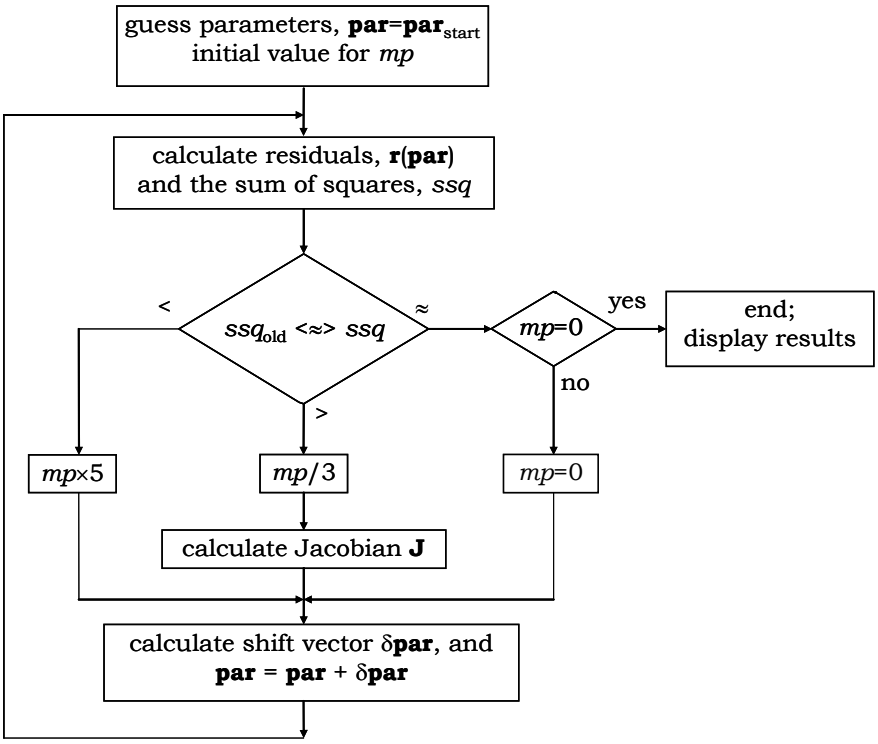

Fig. 3. Flow diagram of the Newton-Gauss-Levenberg/Marquardt (NGL/M) method.

sets. Each example will be used to highlight particular application of the non-linear fitting in chemical equilibria systems.

Example 1: Bi^{3+} and Cl^-

The first example is a spectrophotometric titration of a 2-component system. It is the complexation of Bi^{3+} by chloride ions in aqueous solution. The example is taken from a previous paper [8]. The chloride ion does not act as a base in water and thus the complexation reaction is a pure Lewis acid-base interaction. Bi^{3+} forms octahedral complexes and thus we can expect the step wise complexation of 1 to 6 chlorides. Thus there are 2 components (Bi^{3+}, Cl^-) and altogether 8 species (Bi^{3+}, Cl^-, $BiCl^{2+}$, $BiCl_2^+$, $BiCl_3$, $BiCl_4^-$, $BiCl_5^{2-}$ and $BiCl_6^{3-}$).

The data were generated using formation constants that approximate the ones determined before [8] ($\log\beta_{11}$ = 2.35, $\log\beta_{12}$ = 4.4, $\log\beta_{13}$ = 5.45, $\log\beta_{14}$ = 6.65, $\log\beta_{15}$ = 7.29, $\log\beta_{16}$ = 7.06). The molar absorption spectra of the Bi-species were Gaussian curves which are similar to the spectra determined in [8]. The concentration profiles, the matrix **C**, and spectra, the matrix **A**, are shown in Fig. 4.

Data were generated using Eq. (16) and white noise with a realistic standard deviation of 0.002 was added. The resulting matrix $\mathbf{D}_{\mathbf{meas}}$ is displayed in Fig. 5. The details of the titrations such as initial volumes, added volumes, concentrations *etc.* can be found in the Matlab files supplied by the authors, see end of paper.

Fitting of the data using the correct model and approximate initial guesses for the formation constants results in the formation constants listed in Table 2. As can be seen, all true constants are within the error limits of the fitted values. The numbers in brackets are 2 standard deviations as defined by Eq. (22).

The fitted spectra are essentially correct, they are however influenced by the noise of the measurements. The residuals are normally distributed and the standard deviation, Eq. (23), is correct within the error limits.

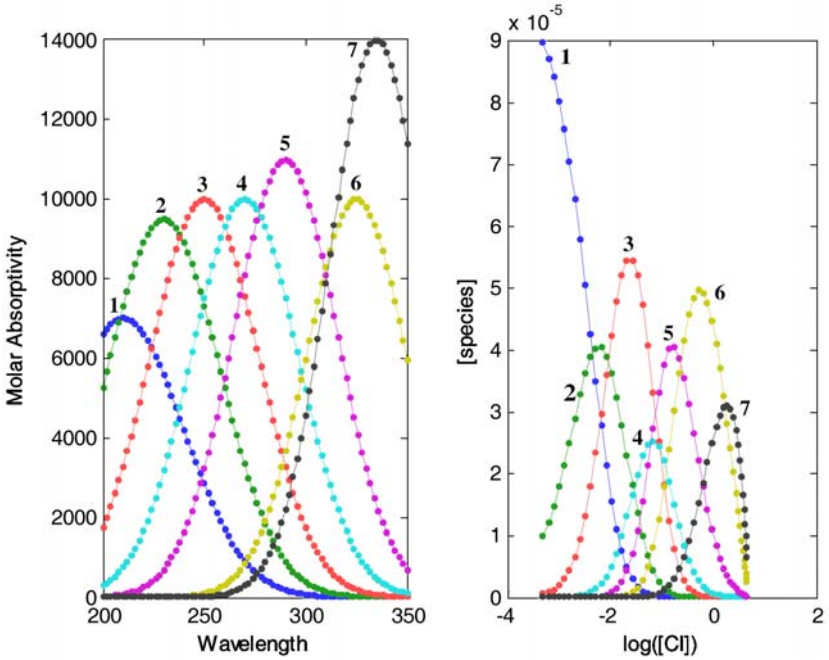

Fig. 5. (a) spectra **A** for the *Bi*-species; (b) concentration profiles **C** of the *Bi*-species as a function of log of the chloride concentration: (1) Bi^{3+}, (2) $BiCl^{2+}$, (3) $BiCl_2^{+}$, (4) $BiCl_3$, (5) $BiCl_4^{-}$, (6) $BiCl_5^{2-}$, (7) $BiCl_6^{3-}$.

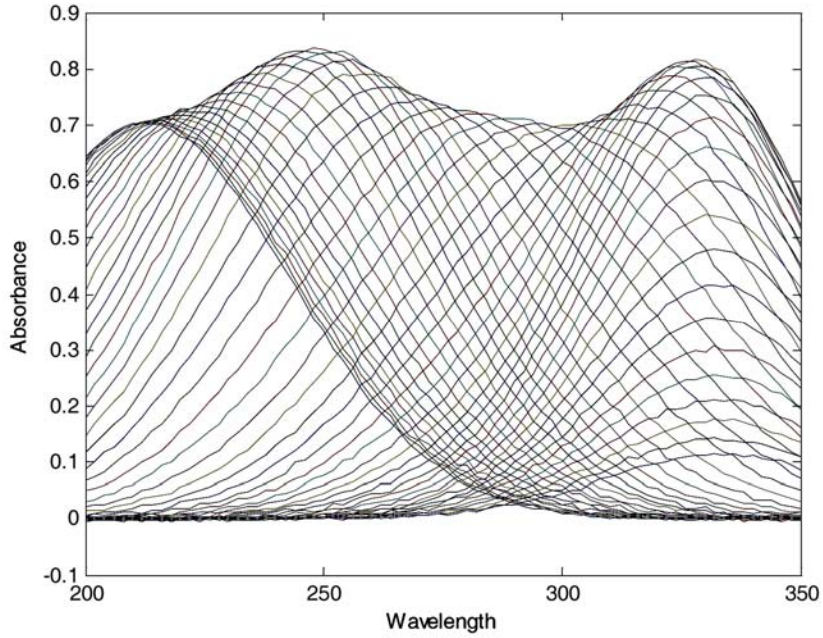

Fig. 5. Series of spectra measured as a function of addition of Cl^-.

Table 2. Fitted Formation Constants and Their True Values for the Bi^{3+}/Cl^{-} System

	Fitted	True
$\log\beta_{11}$	2.37(2)	2.35
$\log\beta_{12}$	4.41(3)	4.4
$\log\beta_{13}$	5.46(6)	5.45
$\log\beta_{14}$	6.65(5)	6.65
$\log\beta_{15}$	7.29(7)	7.29
$\log\beta_{16}$	7.06(9)	7.06
sig_R	0.00198	0.002

Example 2: Cu^{2+} and Edta

The second example is a spectrophotometric titration that investigates the interactions between Cu^{2+} and edta in aqueous solution. Data were taken from a previously published paper [9]. As edta is a Lewis and Brønstedt base we need to include all protonation equilibria, thus we are dealing with a 3-component system, Cu^{2+}, $edta^{4-}$, H^{+}. In addition to the protonation equilibria of edta we also have to deal with protonated and deprotonated complexes. The complete list of all species includes: Cu^{2+}, $edta^{4-}$, H^{+}, $edtaH^{3-}$, $edtaH_2^{2-}$, $edtaH_3^{-}$, $edtaH_4$, $edtaH^{+}$, $edtaH_6^{2+}$, $Cu(edta)^{2-}$, $Cu(edtaH)^{-}$, $Cu(edtaH_2)$, $Cu(edta)H_{-1})^{3-}$ and OH^{-}.

Figure 6 (a) displays the absorption spectra used for the different *Cu*-species and (b) contains their concentration profiles as a function of pH. Figure 7 shows the measurement, the collection of spectra acquired during the titration. As in the proceeding example, the true values for the formation constants are all within the error limits of the fitted values, they are listed in Table 3.

Example 3: Zn^{2+} and Ethylenediamine (en)

The last example is a potentiometric pH titration for the investigation of the complexation of Zn^{2+} with ethylenediamine (*en*). Ethylenediamine is a bidentate ligand which interacts with the octahedral Zn^{2+} ion; it acts also as a base. The following species are formed: Zn^{2+}, *en*, H^{+}, enH^{+}, enH_2^{2+}, $Zn(en)^{2+}$, $Zn(en)_2^{2+}$, $Zn(en)_3^{2+}$ and OH^{-}. The data used for the generation of this measurement are taken from a publishe work [10].

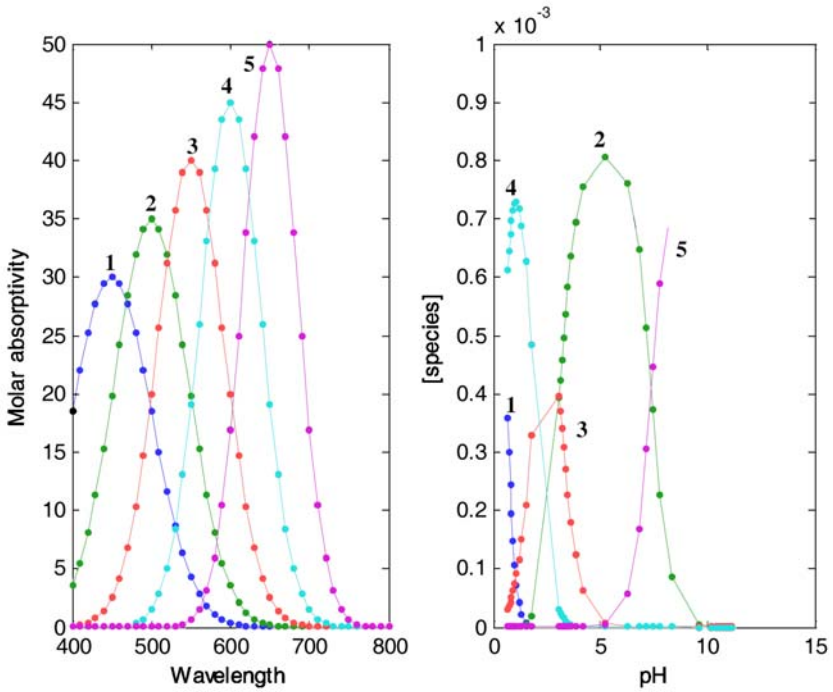

Fig. 6. (a) spectra for the *Cu*-species; (b) concentration profiles of all species (except $H^{+}$and OH^{-}) as a function of pH: (1) Cu^{2+}, (CuL), (3) CuLH, (4) $CuLH_2$, (5) $CuLH_{-1}$.

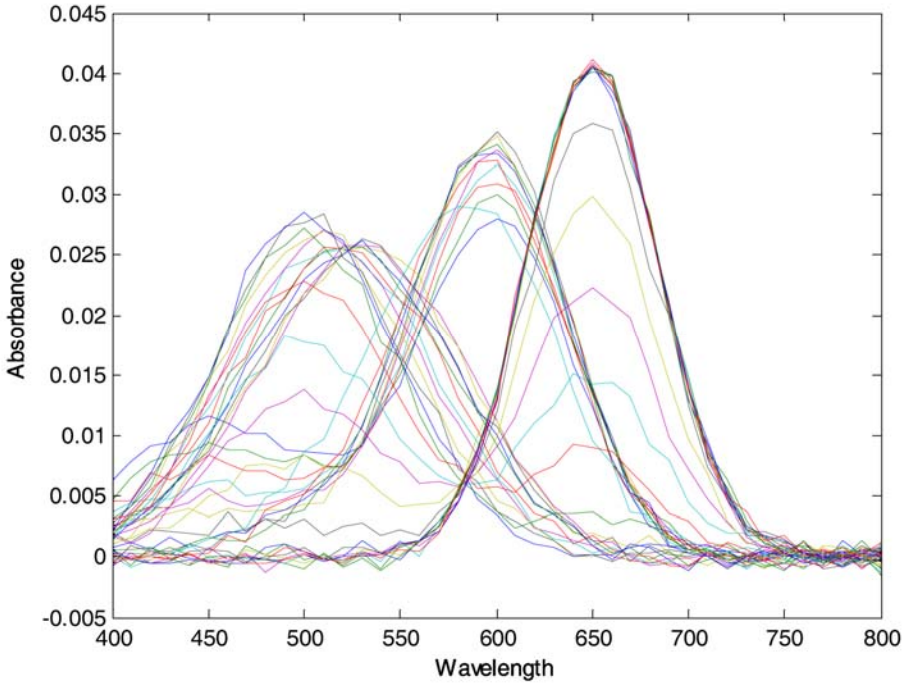

Fig. 7. Series of spectra measured as a function of base addition in the Cu^{2+}/*edta* system.

Table 3. Fitted Formation Constants and Their True Values for the Cu^{2+}/*edta* System

	Fitted	Ttrue
$\log\beta_{110}$	18.81(9)	18.8
$\log\beta_{111}$	21.87(6)	21.9
$\log\beta_{112}$	23.88(7)	23.9
$\log\beta_{11\text{-}1}$	11.3(4)	11.4
sig_R	0.000496	0.0005

Table 4. Fitted Formation Constants and Their True Values for the Zn^{2+}/*en* System

	Fitted	True
$\log\beta_{110}$	5.66(5)	5.69
$\log\beta_{120}$	10.66(4)	10.69
$\log\beta_{130}$	13.1(2)	13
sig_R	0.019	0.02

Fig. 8 shows the data, part (a) is the titration curve, the measured pH (using •-markers and the fitted curve as a line) as a function of added base; (b) contains the concentration profiles for all species. Table 4 contains the fitted parameters and the ones used to generate the data. As before they all are within the error limits.

This data set only includes one titration of a solution that contains *en* and the metal. It is assumed that the protonation constants of *en* are known and can be used for the analysis of this titration.

MATLAB FILES

The collection of Matlab files that were used to generate the above data sets and subsequently to fit the parameters is available from the authors. MM: Marcel.Maeder@newcastle.edu.au; HA: abd@iasbs.ac.ir. These files allow the study of the numerical methods described in this contribution; they also can be adapted for the analysis of other data sets acquired in the laboratory. More explanations on all of the above can be found in reference [4].

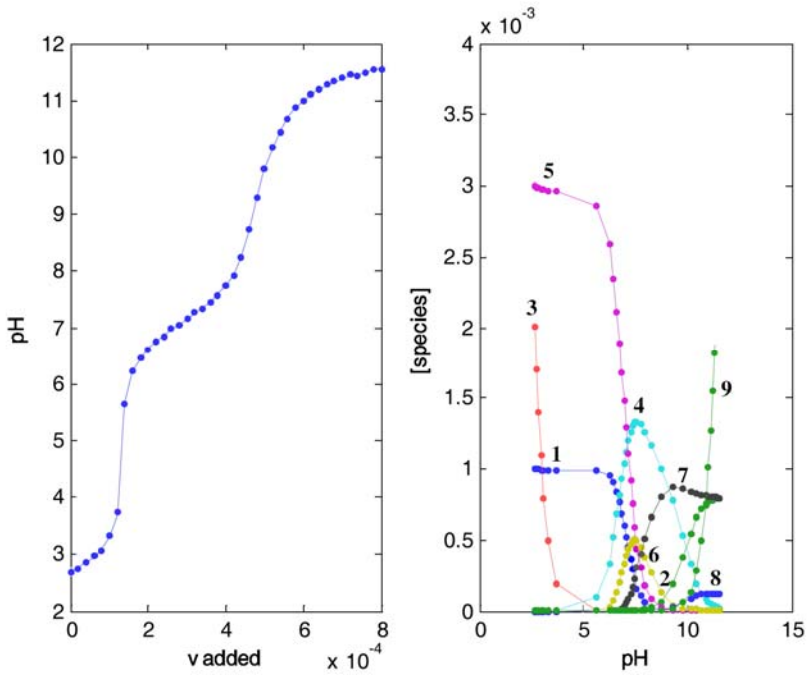

Fig. 8. (a) Potentiometric titration curve; (b) Concentration profiles of all species: (1) Zn^{2+}, (2) L, (3) H, (4) LH, (5) LH_2, (6) ZnL, (7) ZnL_2, (8) ZnL_3, (9) OH.

REFERENCES

[1] A.E. Martell, R.D. Hancock (Eds.), Metal Complexes in Aqueous Solutions, Plenum, New York, 1996.

[2] M.T. Beck, Chemistry of Complex Equilibria, Van Nostrand, London, 1970.

[3] J. Polster, H. Lachmann, Spectrometric Titrations: Analysis of Chemical Equilibria, VCH, Weinheim, 1989.

[4] M. Maeder, Y.-M. Neuhold, Practical Data Analysis in Chemistry, Elsevier, Amsterdam, 2007.

[5] G.H. Golub, F. van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, MD, 1983.

[6] W.H. Press, W.T. Vetterling, S.A. Teukolsky, B.P. Flannery, Numerical Recipes in C, 2nd ed., Cambridge University Press, Cambridge, 1995.

[7] P.R. Bevington, D.K. Robinson, Data Reduction and Error Analysis for the Physical Sciences, 3rd ed., McGraw-Hill, New York, 2002.

[8] J.J. Kankare, Anal. Chem. 42 (1970) 1322.

[9] G. Anderegg, Critical Survey of Stability Constants of EDTA Complexes, Vol. 14, Pergamon Press, Oxford, 1977.

[10] A.E. Martell, R.M. Smith, 6.0 ed., NIST Standard Reference Data, Gaithersburg, 2001.