

Determination of stability constants using genetic algorithms

Margaret K. Hartnett^{a,b,1}, M. Bos^a, W.E. van der Linden^a, Dermot Diamond^{b,*}

^a *Laboratory for Chemical Analysis, Department of Chemical Technology, Twente University of Technology, Enschede, Netherlands*

^b *School of Chemical Sciences, Dublin City University, Dublin, Ireland*

Received 24 January 1995; revised 23 June 1995; accepted 6 July 1995

Abstract

A genetic algorithm (GA)-simplex hybrid approach has been developed for the determination of stability constants using calorimetric and polarographic data obtained from literature sources. The GA determined both the most suitable equilibrium model for the systems studied and the values of the stability constants and the heats of formation for the calorimetric studies. As such, a variable length chromosome format was devised to represent the equilibrium models and stability constants (and heats of formation). The polarographic data were obtained from studies of cadmium chloride and lead with the crown ether dicyclohexyl-18-crown-6. The calorimetric data were obtained from a study of a two step addition reaction of $\text{Hg}(\text{CN})_2$ with thiourea. The stability constants obtained using the GA-simplex hybrid approach compare favourably with the values quoted in the literature.

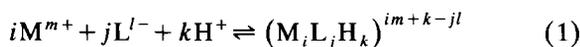
Keywords: Stability constants; Polarography; Calorimetry; Genetic algorithms

1. Introduction

1.1. Stability constants

A stability constant is a thermodynamic quantity which refers to the formation of a complex $\text{M}_i\text{L}_j\text{H}_k$ in which M is usually a metal ion, L refers to a ligand which can be a simple anion or a neutral molecule and H refers to a proton which may be involved in the reaction [1,2].

Mathematically, the stability constant describes the equilibrium



as follows

$$\beta_{ijk}^0 = \frac{\{(\text{M}_i\text{L}_j\text{H}_k)^{im+k-jl}\}}{\{\text{M}^{m+}\}^i \{\text{L}^{l-}\}^j \{\text{H}^+\}^k} \quad (2)$$

where the terms in curly brackets refer to the activities of the species involved in the equilibrium. Since many experimental techniques such as spectroscopy and calorimetry have responses which are dependent on the concentrations of different species rather than their activities, a stoichiometric stability constant may be calculated as follows:

$$\beta_{ijk} = \frac{[(\text{M}_i\text{L}_j\text{H}_k)^{im+k-jl}]}{[\text{M}^{m+}]^i [\text{L}^{l-}]^j [\text{H}^+]^k} \quad (3)$$

* Corresponding author.

¹ Present address: Control Engineering Research Group, Department of Electrical and Electronic Engineering, The Queen's University of Belfast, Belfast, UK.

The terms in brackets in expression (3) refer to the free concentrations of the species involved in the equilibrium as opposed to the total concentrations of the species which are experimental variables of the investigation in question.

If the experiments are performed in media of constant ionic strength, so that the activity coefficients of the species involved do not vary during the experiment then the stoichiometric stability constant β_{ijk} can be converted to the thermodynamic stability constant β_{ijk}° by multiplying it by a constant corresponding to the ratio of the activity coefficients of the metal, ligand and complexed species, raised to the power of their respective charges.

$$\beta_{ijk}^{\circ} = \beta_{ijk} \frac{\gamma_{(M_i L_j H_k)}^{im+k-jl}}{(\gamma_{M^{n+}})^i (\gamma_{L^-})^j (\gamma_{H^+})^k} \quad (4)$$

Otherwise, the stoichiometric coefficient should be corrected for changes during the experiment in the activity coefficients of the species involved in the equilibria. The activity coefficients (γ) of the species involved in this contribution may be calculated by means of the Davies expression [3].

$$\log \gamma_x = -0.5z_x^2 \left(\frac{\sqrt{I}}{1 + \sqrt{I}} - 0.21 \right) \quad (5)$$

In which γ_x is the activity coefficient of species x , z_x is the charge of species x and I is the ionic strength of the medium under investigation.

There are a number of experimental techniques used for the determination of stability constants including potentiometry [4], polarography [5], NMR spectroscopy [6] and calorimetry [7]. Stability constants can be calculated from the experimental data by a number of means of graphical or numerical methods [8,9]. The numerical approaches to the determination of stability constants normally involve the minimisation of a sum of squares or weighted sum of squares parameter. This is achieved by a number of different methods including the Gauss method used by the program SCOGS for pH titration data [10] and the pit mapping technique used by the program DALSF EK for spectroscopic data [11].

A major problem associated with many of these approaches arises from the numerical difficulties posed by high dimensional search spaces associated

with increasingly complex equilibrium models. This is especially true for experimental methods which require the determination of additional parameters such as heats of formation in calorimetry or molar extinction coefficients in spectroscopy, thereby increasing the dimension of the search space. Because of the multi-dimensional nature of the data space involved, the GA was investigated as a means of solving the minimisation problem, because it is known from literature [12] that these algorithms are very suited to dealing with such tasks.

1.2. Polarographic methods

The stability constants of electroactive species which undergo reversible reduction can be determined polarographically by the method of DeFord and Hume [13]. Changes in the half-wave potential and diffusion current of the metal on addition of ligand can be related to the concentration of the metal by means of a term Fo which describes the ratio between the total concentration of metal (in all its complexed and uncomplexed forms) C_M and the concentration of the free metal $[M]$ described by Eq. 6

$$Fo_{(calc)} = \frac{C_M}{[M]} \quad (6)$$

The term Fo can also be related to the experimental variables of the study as described by Eq. 7.

$$Fo_{(expt)} = \exp \left[\frac{nF}{RT} \{E_{1/2S} - E_{1/2C}\} + \ln \left\{ \frac{I_{dS}}{I_{dC}} \right\} \right] \quad (7)$$

where F refers to the Faraday constant, R refers to the gas constant, T refers to temperature in Kelvin and n refers to the number of electrons involved in the electrochemical process. $E_{1/2S}$ and $E_{1/2C}$ refer to the half wave potentials of the metal ion in the absence of ligand and in the presence of ligand, respectively, and I_{dS} and I_{dC} refer to the diffusion limited currents of the metal ion in the absence of ligand and in the presence of ligand, respectively. (The subscripts S and C refer to the simple and complexed forms of the metal ion. This nomenclature is as described in Ref. [13].)

The concentration of free metal ion is related to the total concentration of metal ion by the mass balance expressions involving the stability constants for the system.

$$C_M = \sum_0^i \sum_0^j i \beta_{ij} [M]^i [L]^j \quad (8)$$

$$C_L = \sum_0^i \sum_0^j j \beta_{ij} [M]^i [L]^j \quad (9)$$

The stability constants can thus be determined by a process of minimising the sum of squared errors (SSE) between the experimentally derived value of $Fo(\text{expt})$ (calculated from expression (7)) and the value calculated from the free metal concentrations $Fo(\text{calc})$ (as seen in expression (6)).

$$\begin{aligned} \text{SSE} &= \sum_{y=1}^{\text{num}} \left([Fo(\text{expt})]_y - [Fo(\text{calc})]_y \right)^2 \\ &= \sum_{y=1}^{\text{num}} \left(\exp \left[\frac{nF}{RT} \left\{ E_{1/2S} - (E_{1/2C})_y \right\} \right] \right. \\ &\quad \left. + \ln \left[\frac{I_{ds}}{(I_{dc})_y} \right] - \frac{(C_M)_y}{[M]_y} \right)^2 \quad (10) \end{aligned}$$

The term *num* refers to the number of measurements performed in the experiment and the index *y* refers to the measurement *y* in a particular experiment. Leggett's program POLAG [14] which performs the least squares calculations by means of a Gauss-Newton approach, is one of the more advanced software packages for determining stability constants from polarographic data.

1.3. Calorimetric determination of stability constants

Calorimetric titrations involve the measurement of the temperature of a chemical system during the titration of a species in the system with another species. The results of such experiments describe the temperature of the system as a function of the added titrant (i.e. the results are composed of experimentally observed temperatures recorded for corresponding number of added moles of titrant). The heat evolved from the system as a function of the added

titrant can be determined from the temperature history of the titration using the following expression.

$$\begin{aligned} Q_{(\text{expt})} &= - \left[C_{pr} + V_s \rho_s C_{ps} + V_{t,p} \rho_{t,p} C_{pt,p} \right. \\ &\quad \left. + \left(\frac{\partial C_p}{\partial V} \right) V_{t,p} \right] (T_{\text{start}} - T_p) \quad (11) \end{aligned}$$

In this expression, $Q_{(\text{expt})}$ refers to the heat produced in the reaction vessel, from the start of the titration to any point *p* in the titration history. C_{pr} refers to the heat capacity of the empty reaction vessel, V_s and $V_{t,p}$ refer to the volumes of the initial solution in the reaction vessel and the added titrant respectively. ρ_s and $\rho_{t,p}$ refer to the densities of the initial solution in the reaction vessel and the added titrant respectively. Similarly C_{ps} and $C_{pt,p}$ refer to the specific heat capacities of the initial solution in the reaction vessel and the added titrant respectively. $\left(\frac{\partial C_p}{\partial V} \right) V_{t,p}$ refers to the increase in heat capacity due to the increase in the liquid-wall contact area in the reaction vessel. T_{start} refers to the temperature of the system at the start of the titration and T_p refers to the temperature of the system at any point *p* in the titration history.

Once the heat evolved during the titration has been calculated from the temperature history of the titration, it is then possible to correct these calculated heat values for contributions originating from other sources than the chemical equilibrium being studied. These contributions can arise from a number of sources including, dilution of the titrant and the titrand, heat loss between the reaction vessel and the surroundings, and heat from the mechanical action of stirring of the solution. In these studies, it is assumed that the solvents used for the sample solutions and titrant are the same, or at least that no contraction occurs, otherwise, another source of heat arises from the mixing of two different solvents. As such, the corrected heat values ($Q_{(\text{expt})}^c$) more accurately describe the heat released by the reaction of interest for the corresponding number of moles of added titrant during the titration (provided these contributions can be accurately modelled).

The heat evolved by the complexation reaction can be related to the heats of formation and number of moles of the complexes formed by the reaction. As such, it is possible to calculate a predicted value

for the heat evolved by the complexation reaction ($Q_{(\text{calc})}^c$) by the general expression.

$$Q_{(\text{calc})}^c = \sum_{\text{cmp}=1}^N \Delta H_{\text{cmp}} \text{mol}_{\text{cmp}} \quad (12)$$

In this expression N refers to the number of complexes formed, ΔH_{cmp} refers to the heat of formation of a complex cmp . The term mol_{cmp} refers to the number of moles of the complex concerned present in the reaction vessel which can be calculated from the stability constants and the total concentrations of metal and ligand used for the experiment.

This indicates that in order to predict the heat evolved during the titration one must determine both the stability constants for the complexation reactions and the heats of formation of the complexes, thereby effectively doubling the dimension of the search space. Clearly these parameters cannot be determined from a single temperature measurement, instead the temperature is measured with incremental additions of the titrant during the progress of the titration. The heats of formation and stability constants are generally determined by minimising the sum of squares of the difference between the predicted or calculated values ($Q_{(\text{calc})}^c$) of the heat evolved from the complexation reaction (by means of Eq. 12) and the experimentally derived values ($Q_{(\text{expt})}^c$) depicted in expression (13).

$$\begin{aligned} \text{SSE} &= \sum_{y=1}^{\text{num}} (Q_{(\text{expt})}^c - Q_{(\text{calc})}^c)^2 \\ &= \sum_{y=1}^{\text{num}} \left(Q_{(\text{expt})}^c - \sum_{\text{cmp}=1}^N \Delta H_{\text{cmp}} \text{mol}_{\text{cmp}} \right) \end{aligned} \quad (13)$$

For further details of the theory and application of calorimetric titration to the determination of stability constants the reader is directed to Ref. [15,16].

1.4. Genetic algorithms

Genetic algorithms (GA) are a group of search and optimisation techniques which are based on principles derived from theories of biological evolution. They are members of the class of moderate search techniques. These techniques make fewer assumptions about the nature of the search space than strong

search techniques such as gradient methods and are less computationally demanding than weak techniques such as enumerative methods [12]. This aspect of the GA made it appear a suitable candidate for the high dimensional searches associated with stability constant determination for multiple equilibria.

The operation of the GA involves maintaining a population of P candidate solutions to an optimisation problem (known as chromosomes). The M variables of each candidate solution are encoded in a uniform alphabet on the chromosome as M genes. Variables are commonly encoded in a binary alphabet, either simple binary or Gray's code, yielding a bitstring whose length is determined by the number of variables in the model and the number of bits used to represent each variable.

The degree to which the model represented by a particular chromosome fits the experimental data is described quantitatively by a term called its fitness. Since the determination of stability constants in this study involves the minimisation of a sum of squared errors expression, the fitness of a particular chromosome will be considered to be inversely related to its sum of squared errors (as described in expression (14)).

$$\text{Fitness} = \frac{1}{\text{SSE}} \quad (14)$$

Chromosomes are allowed to reappear in the population in future iterations of the GA with a probability related to its fitness. New areas of the search space are explored by means of two genetic operators, namely crossover and mutation. There are a number of different methods of implementing crossover including multi-point and uniform crossover [12]. In this contribution a single point crossover method was employed. In this form of crossover two chromosomes are selected as mates and the bitstrings representing the model parameters are swapped at positions on the bitstrings chosen at random. During mutation, single bits in the chromosome bitstrings are randomly toggled.

The fitness related reproduction, crossing over and mutation operations are repeated in an iterative cycle as depicted in Fig. 1, until a particular termination criteria is achieved. There are a number of different methods of implementing a GA, for a fuller

discussion of the theory behind GAs and methods of operation the reader is recommended the Refs. [12,17] and [18,19].

The ease of representation of an optimisation problem associated with a GA, appeared suitable for designing a system which was capable of determining both the nature of the equilibrium model and the value of the stability constants involved in this equilibrium. In many software systems, the operator can determine the relative suitability of different equilibrium models by selecting models manually for optimisation and comparing the relevant statistics associated with each model. In this GA approach, a range of different models will be evaluated in parallel removing the need for manual comparison of model performance.

1.5. Genetic representation of the stability constant determination problem

Since the aim of this study was to optimise equilibrium models and stability constants, a variable length chromosome format was decided upon for the population. Fig. 2 graphically depicts the structure of a chromosome used in a simple GA compared with the structure of the chromosome used for this application. In a simple GA the number of parameters to be optimised is generally fixed. As described in Section 1.4 a candidate value for each parameter is usually encoded in a binary form and is called a *gene*. The *genes* are then concatenated together to form a bitstring known as a *chromosome*. As such, a chromosome represents a candidate solution to the particular optimisation problem. In the format used for this study, each *chromosome* was composed of a *table* and a *bitstring*. The *table* described the number of different metals and ligands involved in the candidate equilibrium model, encoded on the chromosome. This information was used to permit the possibility of modelling mixed ligand/metal complexes. The *table* also described the stoichiometric numbers of the metals and ligands in the candidate equilibrium model. As such, the *table* described the candidate model encoded by the chromosome (e.g. ML, ML_2). The representation scheme contained within the *table* was then used to build up a series of equilibria for the lowest order complex ML to the highest order complex specified by the chromosome.

This scheme was then used to calculate the number of stability constants needed to determine the equilibria encoded by the chromosome and hence the total number of parameters needed to be optimised (depending on whether the experimental data used was derived from a polarographic or calorimetric analysis). The *bitstring* encoded the resulting stability constants (and heats of formation for calorimetric studies). The value of each parameter was represented as an integer scaled in the range $0-(2^{16} - 1)$. The resulting string of parameters for each model (the *bitstring*) were dynamically allocated and reallocated according to the breeding processes involved in the GA.

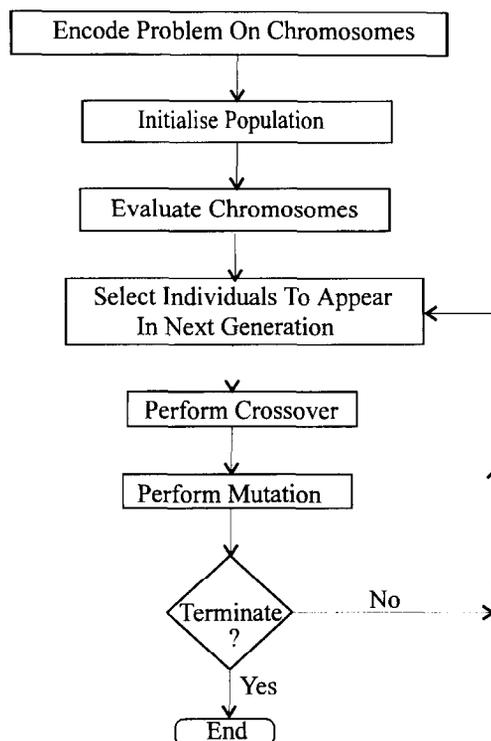


Fig. 1. General processes occurring during the operation of a GA. A population of chromosomes is initialised according to the specific problem encoding scheme. The resulting chromosomes are evaluated according to the objective function of the given problem and are assigned fitness values corresponding to the results of this evaluation. Chromosomes are selected to appear in the next cycle of the GA so that more fit chromosomes are more likely to appear than less fit chromosomes. Chromosomes can then be selected for crossover and mutation. The cycle of evaluation, selection, crossover and mutation continues until a particular termination criterion has been achieved.

The actual configuration for the GAs used for this study is depicted in Table 1. A problem with the implementation of GAs is the low precision with which a solution is determined [12]. While GAs are generally insensitive to local optima in a search space and as such tend to find the approximate area in which a global optimum is likely to exist, the final solutions determined by the GAs when repeated, demonstrate a low degree of precision. In order to “fine-tune” the end solution determined by the GA, the GA was post-hybridised with a high precision local optimisation technique. Post-hybridisation of the GA with a local optimisation technique (in this case simplex optimisation) is intended to implement the complementary aspects of both optimisation approaches. In this case, repeated application of a GA was used to generate members of the population which fit the experimental data well, which were

then used to fill the vertices of the simplex for optimisation refinement.

An alternative approach would be to use a simplex or other local search based on a single implementation of a GA. However if different members of the population of a GA (at termination) were used to initialise the local search, there may be problems associated with the reduced diversity in the population of the GA. Earlier studies in this regard were unfortunately inconclusive [22].

1.6. Implementation of objective function

In the previous sections discussing polarography and calorimetry, it was shown that stability constants determined by these techniques were calculated by means of optimising a sum of squared errors expression given by Eqs. 10 and 12 respectively. It can be

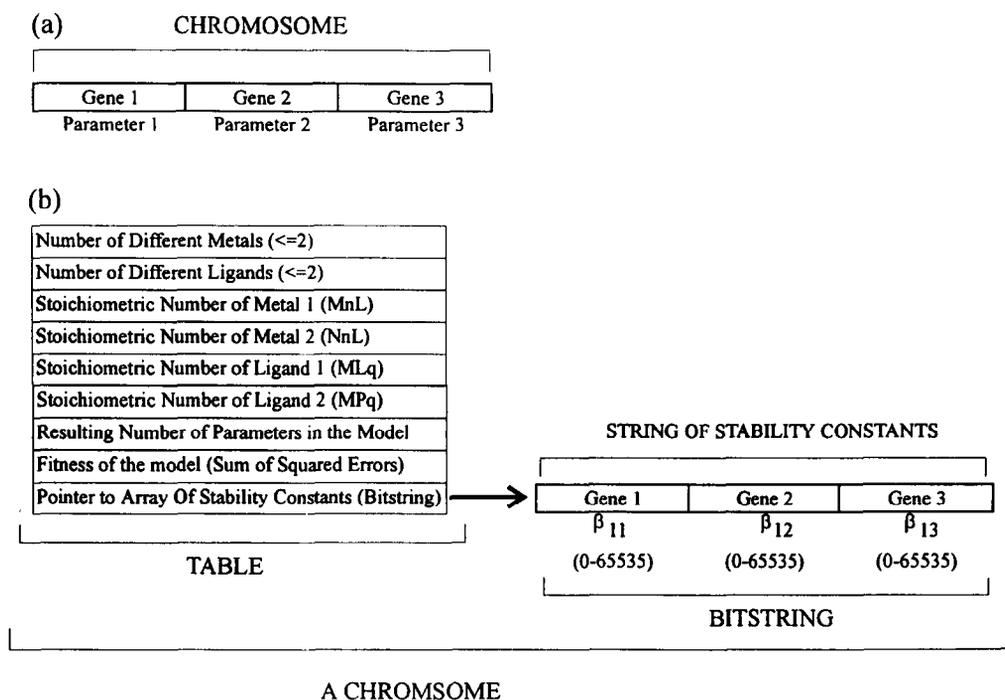


Fig. 2. (a) Representation of a candidate solution to an optimisation problem with a conventional GA. The number of parameters to be optimised is fixed. A CHROMOSOME is composed of a string of GENES which encode the parameters of the optimisation problem in binary form. (b) Representation of a candidate solution as applied to the stability constant determination problem. The number of parameters to be optimised is variable. A CHROMOSOME is composed of a TABLE and a BITSTRING. The TABLE entries describe the candidate model and the BITSTRING encodes the resulting stability constants in a binary form. While the figure shows a chromosome in which three parameters are coded as 16 bits each, the GA is designed to use variable length chromosomes to encode different order equilibrium models, as such the population would be composed of chromosomes with variable length bitstrings.

seen that both expression (10) and (12) involve the equilibrium concentrations of the species described by the equilibrium model. The equilibrium concentrations of the species can be calculated from the total metal and ligand concentrations (which are experimental variables) and from the proposed equilibrium model and stability constants, using the mass balance expressions (8) and (9). As such, it can be seen that the calculation of a fitness measure for a chromosome is an indirect procedure which involves using the chromosome encoded model and experimental data describing total metal and ligand concentrations to calculate the equilibrium concentrations of the species described by the model. This is achieved by a separate optimisation routine which minimises the difference between the total metal and ligand concentrations predicted from the mass balance expressions and the total metal and ligand concentrations used experimentally. These equilibrium concentrations are then used to predict a value of F_o or corrected heat value respectively for polarographic data or calorimetric data. The F_o or heat values are

then compared with the experimentally determined values for these parameters to generate a fitness measure which is inversely related to the sum squared error calculated from Eqs. 10 and 13 for the chromosome. As such, the separate optimisation procedure does not optimise any of the stability constant or heat of formation parameters. Rather, it optimises the values of the equilibrium concentrations of the species described by the particular equilibrium model encoded on a chromosome.

1.7. Parameter standard deviations

A major advantage of computational methods of determination of stability constants over graphical methods is the ability to estimate the precision of the model parameters determined by the optimisation procedure.

The principle of the GA approach involves the minimisation of a sum of squared errors value, calculated as the difference between an experimental ob-

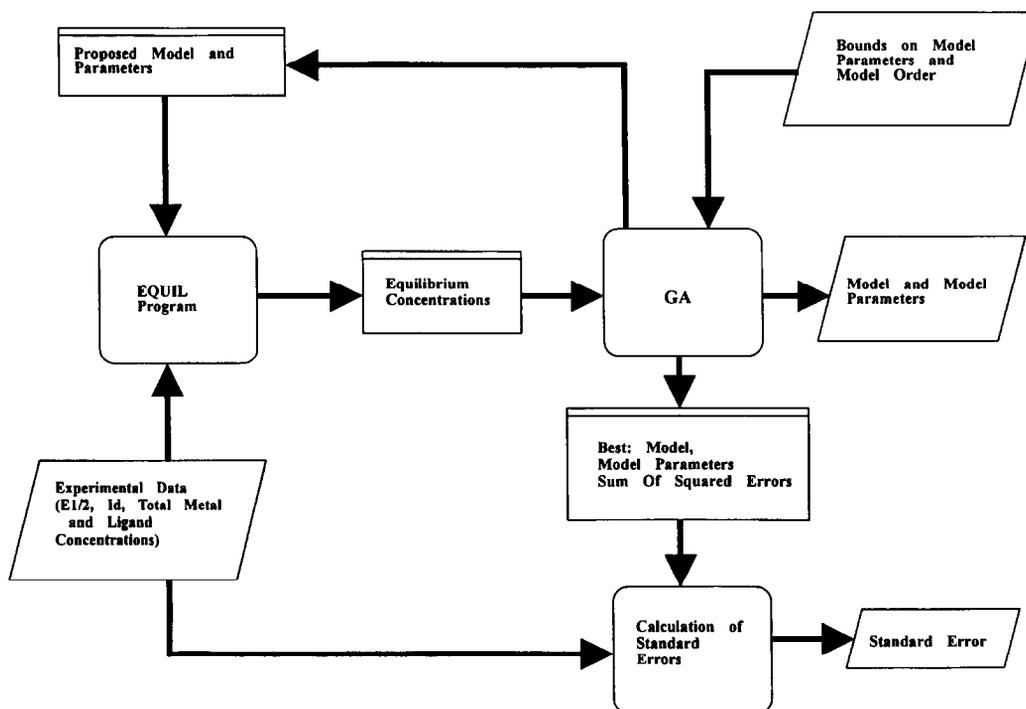


Fig. 3. Dataflow diagram for the GA and associated error handling software.

servation and a corresponding value calculated from the equilibrium model as follows:

$$\text{SSE} = \sum_{i=1}^m W_i \left[y_{i(\text{expt})} - f(\alpha_1 \dots \alpha_n, x_{i,1} \dots x_{i,k}) \right]^2 \quad (15)$$

Where W_i represents the weighting of each experimental point (assumed to be unity in this case), m represents the number of observations, $y_{i(\text{expt})}$ represents the experimental observation (Fo for polarography and heat for calorimetry) and $f(\alpha_1 \dots \alpha_n, x_{i,1} \dots x_{i,k})$ represents the function calculating a value corresponding to the experimental value. In this case f is a function of x_l ($l = 1$ to k) independent variables (e.g. total metal and ligand concentrations) and α_j ($j = 1$ to n) parameters (e.g. stability constants, heats of formation).

The principle of propagation of errors described by Deming [23] can be used to determine the standard errors of the function parameters α_j by means of the linear term in the Taylor's expansion of the function f . The variance of each function parameter can be determined from the diagonal elements of the matrix $\mathbf{B}\sigma^2$. \mathbf{B} is calculated as $(\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1}$ where \mathbf{J} is the Jacobian matrix of the function depicted as follows

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1}{\partial \alpha_1} & \dots & \frac{\partial f_1}{\partial \alpha_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial \alpha_1} & \dots & \frac{\partial f_m}{\partial \alpha_n} \end{bmatrix} \quad (16)$$

Table 1
General configuration of genetic algorithms used for this study

Population	200–300 chromosomes
Scaling	Rank scaling [20]
Rank scaling upper limit	1.1
Selection	Stochastic remainder [17]
Crossover	Single point [17]
Crossover restriction	Chromosome mates must encode the same number of parameters Hamming distance restriction [21]
Cross-over rate	0.65
Mutation	Single point
Mutation rate	1.0/(population size)

Table 2

Polarographic data used for the determination of the stability constants of the cadmium(II) chloride system

$[\text{Cl}^-]_{\text{T}}$ (M)	$-E_p$ (V vs. Ag/AgCl)	I_p (μA)
0	0.5885	7.10
0.05	0.5989	7.25
0.1	0.6057	7.37
0.2	0.6151	7.37
0.4	0.6277	7.37
0.6	0.6365	7.40
0.80	0.6435	7.40
1.00	0.6493	7.40

E_p and I_p represent the peak potential and peak current respectively, $[\text{Cl}^-]_{\text{T}}$ represents the total concentration of the species in square brackets.

σ^2 represents the variance between the experimental observation and the calculated value. The partial derivatives contained in the Jacobian were calculated numerically in this study by means of Ridder's method of polynomial extrapolation [24] and a singular value decomposition [24] approach was used to invert the matrix \mathbf{B} .

For further details of this approach, the reader is directed to Refs. [25–27]. A summary of the operation of the GA and the associated error handling software is depicted in the dataflow diagram for the in Fig. 3. Starting with the experimental data and an expected range of models and values for stability constants, a population of chromosomes is established in the GA using the encoding scheme depicted in Fig. 2b. A candidate model and parameters encoded on a chromosome together with the total metal and ligand concentrations (contained in the experimental data) are used to calculate equilibrium concentrations by means of the EQUIL program. The equilibrium concentrations are then used by the GA to evaluate the fitness of the candidate chromosome. The GA evaluates all the chromosomes in the population the results of which are used to select chromosomes to appear in the next generation of the GA. The chromosomes are subjected to crossover and mutation as in Fig. 1. The cycle of evaluation, selection, crossover and mutation continues until the GA terminates. At this point the model and parameters which best fit the experimental data are used with the experimental data to calculate the standard errors of the parameters.

2. Experimental

The GA software was written in ANSI style C running on an Atari ATW-800 transputer workstation using the Helios operating system. The equilibrium concentrations of the different species present in a given equilibrium were calculated by means of the program EQUIL [28] which was merged with the GA.

3. Results

3.1. Determination of stability constants from polarographic experiments

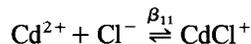
3.1.1. Complexation of cadmium chloride in a perchlorate medium

The ability of the GA to determine stability constants from data obtained from polarographic experiments was tested initially with data published by Heath and Hefter [29] concerning a cadmium(II) chloride system studied by means of differential pulse polarography. It involved the measurement of the peak potential and peak current from the polarographic reduction of cadmium(II) in 1.0 M sodium perchlorate with varying concentrations of sodium chloride. The total concentration of cadmium(II) was 4.0×10^{-5} M and the chloride concentration varied over the range 0.05 to 1.00 M. The raw data used for this study are given in Table 2.

In order to determine the stability constants from these data the original DeFord and Hume expressions (7) for data obtained from linear potential sweep (d.c.) polarography were altered by Heath and Hefter [29] for the differential pulse (d.p.) technique, by replacing the half-wave potential and diffusion

currents with the peak potential and peak current respectively.

In this system, three different complexes are predicted to exist, namely CdCl^+ , CdCl_2 and CdCl_3^- in concentrations determined by the stability constants, β_{11} , β_{12} and β_{13} of the equilibria:



This system was also studied by Leggett [14] with the POLAG program and by Laouenan and Suet [30] with the MICMAC program. The estimates of the stability constants provided by these different studies are given in Table 3.

It can be seen from Table 3 that there are some differences between the values of the stability constants calculated by Heath and Hefter [29] and the other studies. It can also be seen that there is much closer agreement between the values of the stability constants determined by Leggett [14] and Laouenan and Suet [30]. This was most likely to have been caused by Heath and Hefter's [29] use of a graphical method to calculate the stability constants from their experimental data whereas Leggett [14] and Laouenan and Suet [30] used more sophisticated numerical techniques. Bond [31] has shown how there can be a wide variation between the values of stability constants determined from the same graphical approach. Part of this problem arises from personal bias in the extrapolation procedure used to calculate values of stability constants from the graphs. Another problem with this approach is that stability constants are

Table 3

Values of the stability constants for cadmium chloride complexation determined previously from literature sources based on data acquired from a differential pulse polarography experiment performed by Heath and Hefter

Parameter	Literature values		
	Heath and Hefter [29]	Leggett [14]	Laouenan and Suet [30]
Log β_{11}	1.352	1.329	1.328
Log β_{12}	1.748	1.736	1.735
Log β_{13}	1.544	1.514	1.511

calculated in a propagative fashion, leading to an accumulation of errors.

A GA with a configuration similar to that in Table 1 was used for this study except for employing a population containing 200 chromosomes and a rank prescaling constant of 1.3, and was allowed to iterate thorough 100 cycles before termination. The population of chromosomes was initialised to search for a suitable model for the complexation of the cadmium chloride from a range of possible models

(1) $M + L \rightarrow ML$ with stability constant β_{11} describing the equilibrium $M + L \rightleftharpoons ML$;

(2) $M + 2L \rightarrow ML_2$ with stability constants β_{11} describing the equilibrium $M + L \rightleftharpoons ML$ and β_{12} describing the equilibrium $M + 2L \rightleftharpoons ML_2$;

(3) $M + 3L \rightarrow ML_3$ with stability constants β_{11} describing the equilibrium $M + L \rightleftharpoons ML$, β_{12} describing the equilibrium $M + 2L \rightleftharpoons ML_2$ and β_{13} describing the equilibrium $M + 3L \rightleftharpoons ML_3$.

The GA was also initialised to search for stability constants in the range $0 \leq \log \beta \leq 2$. The resulting GA was repeated five times and in each case converged to the correct model for the cadmium chloride complexation (in terms of the stoichiometric coefficients of the metal and ligand involved in the equilibria). Table 4 depicts the statistics concerning the values of the stability constants determined from the repetitions of the GA.

These results show that the first stability constant predicted by the GA agrees more closely with that predicted by the literature sources than the other stability constants of the model. It can also be seen that the first stability constant was more precisely determined on repetition of the GA than the other

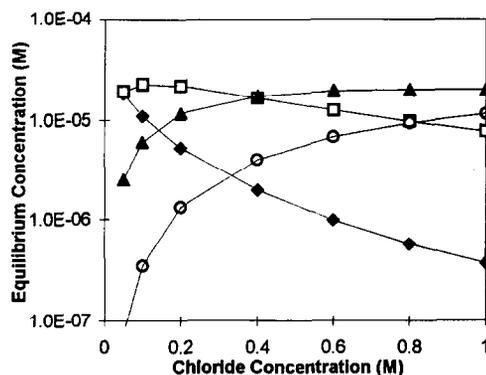


Fig. 4. Distribution of cadmium between its free and complexed forms as a function of chloride concentration for a total cadmium concentration of 4.0×10^{-5} M as calculated from the stability constants determined by the GA-simplex hybrid (the values of these stability constants are depicted in Table 4). Key: \blacklozenge = [Cd]; \square = [CdCl]; \blacktriangle = [CdCl₃].

stability constants. The stability constants determined from the repetitions of the GA were used to provide the vertices of a simplex to refine the model determined by the GA and the standard deviations of the parameters in the refined model were then determined by means of the techniques discussed in Section 1.7. The resulting parameters and their standard deviations can also be seen in Table 4.

It can be seen from Table 4 that there is a particular improvement in the agreement between the literature quoted values and GA-simplex determined values of the stability constants encoding higher order complexes. It can also be seen that the standard deviation of the stability constant describing CdCl₃⁻ has the largest value. This may be explained by

Table 4

The mean values of the stability constants of cadmium chloride complexation determined from the differential pulse polarographic data of Heath and Hefter [29] by multiple repetitions of a GA with 200 chromosomes in its population configured as described in Table 1 (except for a rank prescaling constant of 1.3) to search for stability constants whose values lay in the range $0 \leq \log \beta \leq 2$. The table also shows the refined values of these stability constants following post-hybridisation with a simplex method

Parameter	Mean value of the parameter determined after repetition of The GA ^a	Parameter value after refinement of a GA model using a Simplex method ^b
Log β_{11}	1.301 (0.025)	1.322 (0.003)
Log β_{12}	1.804 (0.059)	1.742 (0.008)
Log β_{13}	1.290 (0.224)	1.502 (0.015)

^a The term in parentheses refers to the standard deviation of the parameters determined from multiple repetitions of the GA.

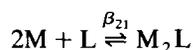
^b The term in parentheses refers to the standard deviations of the parameters determined by the method described in Section 1.7.

examining the distribution plot for the cadmium complexes depicted in Fig. 4, where it can be seen that the CdCl_3^- complex remains at a very low concentration relative to the other cadmium containing species until the later stages of the titration (i.e. chloride concentrations $> 0.6 \text{ M}$) and as such it contributes very little to the estimate of the total concentration of cadmium (as determined from expression 8). As discussed in the introduction, the estimated value for the total concentration of cadmium is in turn used for the calculation of an F_0 value (see expression 6) which is then used in the objective function in the minimisation process of the GA. Since the CdCl_3^- complex contributes little to the variable used in the objective function until the later stages of the titration, whereas the CdCl^+ and CdCl_2 complexes contribute more to the total cadmium concentration estimate for the duration of the experiment (except for the CdCl^+ complex at the chloride concentrations $> 0.8 \text{ M}$), it is more likely that the stability constants associated with the CdCl^+ and CdCl_2 complexes will be determined more precisely than the stability constant associated with the CdCl_3^- complex.

3.1.2. Complexation of lead by the crown ether dicyclohexyl-18-crown-6 in 0.1 M methanol

This study concerned the determination of the equilibrium model and stability constants for the complexation of lead by dicyclohexyl-18-crown-6 in methanol from sampled d.c. polarography data originally investigated by Chen et al. [32] The total concentration of lead(II) varied from 0.5 mM to 0.455 mM and the concentration of the crown ether varied from 0.98 mM to 4.5 mM.

The raw experimental data for this study are depicted in Table 5. The study by Chen et al. involved the use of the POLAG program which suggested that two species could exist, described by the following equilibria:



The values of the stability constants determined by the POLAG program for the experimental data

Table 5
Polarographic data for the system Pb^{2+} –dicyclohexyl-18-crown-6–0.1 M TBAP–methanol

[Ligand] _T (10^{-3} M)	[Lead (II)] _T (10^{-3} M)	$E_{1/2}$ (V)	I_d (μA)
–	0.500	–0.3038	1.290
0.980	0.490	–0.4300	1.211
1.456	0.485	–0.4377	1.217
1.923	0.480	–0.4412	1.211
2.381	0.476	–0.4442	1.202
2.830	0.472	–0.4470	1.201
3.704	0.463	–0.4511	1.181
4.128	0.459	–0.4532	1.181
4.546	0.455	–0.4551	1.178

[_T] represents the total concentration of the species in the square brackets.

were $\log \beta_{11} = 7.46 \pm 0.01$ and $\log \beta_{21} = 14.2 \pm 0.1$.

The initial GA used for this study had a population of 100 chromosomes and 60 chromosomes from the population were allowed to cross. The population of chromosomes was initialised to search for a suitable model for the complexation of the lead by the crown ether from a range of possible models:

(1) $\text{M} + \text{L} \rightarrow \text{ML}$ with stability constant β_{11} describing the equilibrium $\text{M} + \text{L} \rightleftharpoons \text{ML}$.

(2) $2\text{M} + \text{L} \rightarrow \text{M}_2\text{L}$ with stability constants β_{11} describing the equilibrium $\text{M} + \text{L} \rightleftharpoons$ and β_{21} describing the equilibrium $2\text{M} + \text{L} \rightleftharpoons \text{M}_2\text{L}$.

(3) $\text{M} + 2\text{L} \rightarrow \text{ML}_2$ with stability constants β_{11} describing the equilibrium $\text{M} + \text{L} \rightleftharpoons \text{ML}$ and β_{12} describing the equilibrium $\text{M} + 2\text{L} \rightleftharpoons \text{ML}_2$.

(4) $2\text{M} + 2\text{L} \rightarrow \text{M}_2\text{L}_2$ with stability constants β_{11} describing the equilibrium $\text{M} + \text{L} \rightleftharpoons \text{ML}$, β_{12} describing the equilibrium $\text{M} + 2\text{L} \rightleftharpoons \text{ML}_2$, β_{21} describing the equilibrium $2\text{M} + \text{L} \rightleftharpoons \text{M}_2\text{L}$ and β_{22} describing the equilibrium $2\text{M} + 2\text{L} \rightleftharpoons \text{M}_2\text{L}_2$.

The GA was also restricted to search for stability constants in the range $7 \leq \log \beta \leq 18$. Initial studies with the GA yielded, $2\text{M} + \text{L} \rightarrow \text{M}_2\text{L}$ and $2\text{M} + 2\text{L} \rightarrow \text{M}_2\text{L}_2$ models with stability constants β_{11} and β_{12} (when the model encoded an ML_2 complex) in the approximate range of 7.3 to 7.5 and β_{21} and β_{22} (when the model encoded an M_2L_2 complex) in the approximate range 13.7 to 15.7. Based on this initial study the range of models to be searched by the GA was reduced to $2\text{M} + \text{L} \rightarrow \text{M}_2\text{L}$ and $2\text{M} + 2\text{L} \rightarrow$

Table 6

Values of the stability constants determined for the complexation of lead(II) by dicyclohexyl-18-crown-6 in 0.1 M methanol by a GA configured as described in the text a GA-simplex hybrid and the POLAG program

Optimisation technique	β_{11}	β_{21}	Error
GA	7.47	14.04	1.0×10^{-4} (SSRE) ^a
GA – Simplex hybrid	7.459 (0.08)	14.147 (0.66)	8.4×10^{-5} (SSRE)
POLAG program [29]	7.46 (0.01)	14.3 (0.1)	1.9×10^{-6} (SSE)

The terms in brackets beside the stability constants determined by the GA-simplex hybrid refer to the standard deviations of the parameters determined in the manner described in Section 1.7.

^a SSRE corresponds to sum of squared residual errors

$$\text{SSRE} = \sum_{y=1}^{\text{num}} \left(\frac{[Fo_{(\text{expt})}]_y - [Fo_{(\text{calc})}]_y}{[Fo_{(\text{expt})}]_y} \right)^2$$

whereas SSE is a sum of squared errors

$$\text{SSE} = \sum_{y=1}^{\text{num}} ([Fo_{(\text{calc})}]_y - [Fo_{(\text{expt})}]_y)^2$$

M_2L_2 , and the stability constant range to be searched was reduced to $7 \leq \log \beta \leq 16$. The newly configured GA yielded $2M + L \rightarrow M_2L$ and $2M + 2L \rightarrow M_2L_2$ models with very little difference in their SSE. As such it was decided to use members from the population yielding the simpler model to con-

struct the vertices of a simplex for further refinement of the model.

Table 6 depicts the values of the stability constants determined for the $2M + L \rightarrow M_2L$ model determined by the GA and GA-simplex hybrid.

It can be seen from Table 6 that there is a close

Table 7

Calorimetric data for the two step addition reaction of thiourea to $Hg(CN)_2$

Time (s)	Volume of added ligand (ml)	[Thiourea] _T (M)	[Hg(CN) ₂] _T (M)	Corrected heat Q_c (cal)
54.06	0.307493	0.00436	0.030008	-0.44
114.06	0.648773	0.009167	0.029906	-0.91
174.06	0.990053	0.013942	0.029805	-1.39
234.06	1.331333	0.018685	0.029704	-1.85
294.06	1.672613	0.023396	0.029605	-2.31
354.06	2.013893	0.028075	0.029506	-2.76
414.06	2.355173	0.032723	0.029407	-3.2
474.06	2.696453	0.03734	0.02931	-3.61
534.06	3.037733	0.041927	0.029213	-4.02
594.06	3.379013	0.046484	0.029116	-4.42
654.06	3.720293	0.05101	0.02902	-4.81
714.06	4.061573	0.055507	0.028925	-5.18
774.06	4.402853	0.059974	0.028831	-5.54
834.06	4.744133	0.064412	0.028737	-5.89
894.06	5.085413	0.068822	0.028643	-6.23
954.06	5.426693	0.073202	0.02855	-6.56
1014.06	5.767973	0.077555	0.028458	-6.88
1074.06	6.109253	0.08188	0.028367	-7.18

The titrant delivery rate used for calculation of the volume of thiourea added, was 0.005688 ml/s.

Table 8

Comparison of stability constants and heats of formation determined from a calorimetric study of the two step addition reaction of thiourea with $\text{Hg}(\text{CN})_2$ from Ref. [30] by GA and GA-simplex hybrid with the values for these variables determined in the reference

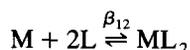
	Literature	GA	GA-simplex hybrid
Log β_{11}	2.074	1.411	2.124 (0.132)
Log β_{12}	2.644	2.186	2.687 (0.105)
$\Delta H(\text{ML})$ kcal/mol	-1.279	-2.303	-1.248 (0.08)
$\Delta H(\text{ML}_2)$ kcal/mol	-9.957	-7.231	-10.113 (0.57)
SSE	0.00018	0.000316	0.000215

The terms in brackets beside the stability constants determined by the GA-simplex hybrid refer to the standard deviations of the parameters determined in the manner described in Section 1.7.

agreement between the stability constants determined by the POLAG program and the GA-simplex hybrid.

3.2. Determination of stability constants from calorimetric experiments

The ability of the GA to determine stability constants from calorimetric data was tested with data published by Eatough et al. [33] concerning the two step addition reaction of $\text{Hg}(\text{CN})_2$ with thiourea whose resulting equilibria are described as follows:



where M corresponds to $\text{Hg}(\text{CN})_2$ and L corresponds to thiourea. The experiment involved the measurement of the heat released on the titration of 99.99 ml of 0.03010 F $\text{Hg}(\text{CN})_2$ with 1.422 F thiourea. The raw experimental data used for this study is given in Table 7.

The stability constants calculated from the resulting data by Eatough et al. [33] had the values $\log \beta_{11} = 2.074$, $\log \beta_{12} = 2.644$ and the heats of formation for the resulting complexes had the values ΔH for the ML complex = -1.279 kcal/mol and ΔH for the ML_2 complex = -9.957 kcal/mol. The GA for this study was initialised to search for stability constants in the range $0 \leq \log \beta \leq 3$ and heat of formation values in the range $-1.0 \text{ kcal/mole} \leq \Delta H \leq -10.0 \text{ kcal/mole}$. Repetitions of the GA yielded either ML_2 or ML_3 models with little difference in their SSE values but widely varying ΔH values. Members of the population from the GA

which converged to the simplest model (i.e. ML_2) were used to construct the vertices of a simplex for refinement of the model. The results of this study are depicted in Table 8.

As can be seen from the table there is an obvious difference between the parameters of the models predicted by the different techniques with relatively large standard deviations for the stability constant and the heat of formation for the ML_2 complex. However on examination of the correlation matrix for the stability constants and heats of formation determined by the GA-simplex hybrid as depicted in Table 9, it can be seen that the first stability constant is very highly correlated with the second stability constant. In addition the first heat of formation of the ML complex and the second stability constant are also highly correlated with the heat of formation for the ML complex.

This high degree of correlation between some of the variables may have caused some of the problems encountered by the GA by making it difficult to find a unique solution for the modelling problem in the highly correlated variables. In addition to this it can be seen from Fig. 5 that the ML_2 complex has a very low concentration relative to the concentration of the ML complex until a thiourea concentration of about

Table 9

Correlation matrix for the stability constants and heats of formation for the two-step addition reaction of $\text{Hg}(\text{CN})_2$ with thiourea

	Log β_{11}	Log β_{12}	$\Delta H(\text{ML})$	$\Delta H(\text{ML}_2)$
Log β_{11}	1.000	0.987	0.999	-0.878
Log β_{12}	0.987	1.000	0.990	-0.790
$\Delta H(\text{ML})$	0.999	0.990	1.000	-0.868
$\Delta H(\text{ML}_2)$	-0.878	-0.790	-0.868	1.000

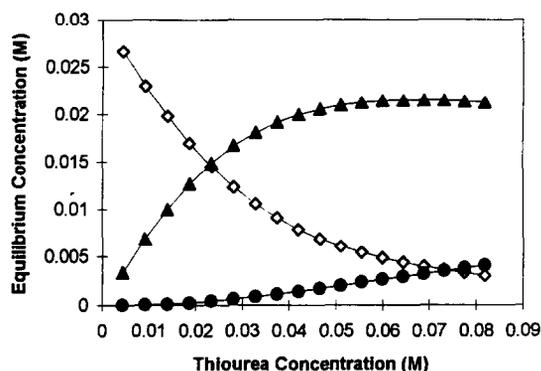


Fig. 5. Distribution of $\text{Hg}(\text{CN})_2$ between its free and complexed forms as a function of thiourea concentration for a total $\text{Hg}(\text{CN})_2$ concentration varying between 0.03 M and 0.028 M as calculated from the stability constants determined by the GA-simplex hybrid (the values of these stability constants are depicted in Table 5). Key: $\diamond = [\text{M}]$; $\triangle = [\text{ML}]$; $\bullet = [\text{ML}_2]$.

0.03 M is reached. As such it may have been more difficult to estimate the heat contribution from the ML_2 complex as precisely as the heat contribution from the ML complex as is demonstrated by the larger standard deviation of the former in Table 8.

4. Conclusion

From the discussion of the results it is clear that there are grounds for a great deal of further improvement in the present GA. The use of an unweighted sum of squares expression for the objective function for this study makes the assumption that all the experimental measurements are performed with equal precision. This may not be a valid assumption and it is suggested that any future studies with this approach might consider this. Similarly, the direct comparison of different order models purely on the basis of differences in their SSEs is clearly not adequate. A more useful approach might be to perform ANOVA on the models. *F*-Distributions for each model could readily be generated and the position of the ratio of the variance due to regression and about regression could be used to weight the SSE of a given chromosome. A problem with this approach however, is that it would require the existence of replicate measurements. If such data were not available, an alternative approach would be to use a cross validation method of selecting models on the basis

their PRESS (predictive residual sum of squares) determined by a leave one out method [34]. In addition, further studies could be performed into improving the GA configuration itself. The exploration aspect of the GA is limited to the parameter space and does not use mutation or crossover to explore different model orders. There are a range of different methods of exploring variable dimension spaces (or variable order models) including the messy GA [35], the SAGA cross [36] and genetic programming [37]. Other selection schemes such as tournament selection [38] could be investigated or the positional bias of the single point crossover (which is especially a problem with highly correlated variables) could be reduced with a uniform crossover.

Having discussed the weaknesses with this approach, it can be seen from the preceding discussions that this approach represents a potentially very powerful method of determining both equilibrium parameters (e.g. stability constants) and complexation models. This contribution has shown that it is possible to use a GA to automatically evaluate different models in parallel rather than the more conventional manual model comparison. The GA as described has clearly demonstrated the ability to determine complexation models and stability constants and heats of formation in close agreement with those determined in the literature.

5. Symbol Conventions For Expressions

M	Metal involved in a complex
L	Ligand involved in a complex
H	Proton involved in a complex
$m +$	Charge of the free, uncomplexed metal ion
$l -$	Charge of the free, uncomplexed ligand ion
i	Stoichiometric number of the metal in a complex
j	Stoichiometric number of the ligand in a complex
k	Stoichiometric number of the proton in a complex
β_{ijk}^0	Stability constant of the complex $(\text{M}_i\text{L}_j\text{H}_k)^{i m + k - j l}$
β_{ijk}	Stoichiometric stability constant of the complex $(\text{M}_i\text{L}_j\text{H}_k)^{i m + k - j l}$

$\gamma_{M^{m+}}$	Activity coefficient of the free uncomplexed metal ion	V_s	Volume of the initial solution in the reaction vessel
γ_{L^-}	Activity coefficient of the free uncomplexed ligand ion	$V_{t,p}$	Volume of the added titrant
γ_{H^+}	Activity coefficient of the free uncomplexed protons	ρ_s	Density of the initial solution in the reaction vessel
$\gamma_{(M_iL_jH_k)_{im+k-jl}}$	Activity coefficient of the complex	$\rho_{t,p}$	Density of the added titrant
γ_x	Activity coefficient of the species x	C_{ps}	specific heat capacity of the initial solution in the reaction vessel
Z_x	Charge of the species x	$C_{pt,p}$	specific heat capacity of the added titrant respectively.
I	Ionic strength of a medium	T_{start}	Temperature at the start of a calorimetric titration
C_M	Total concentration of metal (in all complexed and uncomplexed forms)	T_p	Temperature at any point p in a calorimetric titration
C_L	Total concentration of ligand (in all complexed and uncomplexed forms)	p	Measurement p in a calorimetric titration
[M]	Equilibrium concentration of free uncomplexed metal	$Q_{(expt)}^c$	Heat calculated from the experimental measurements of the temperature history of a calorimetric titration corrected for contributions arising from other sources than the chemical equilibrium being studied (see Eq. 11)
$Fo_{(calc)}$	Calculated ratio between the total concentration of metal (C_M) and the equilibrium concentration of free uncomplexed metal [M] (see Eq. 6)	N	Total number of complexes formed
n	Number of electrons involved in an electrochemical process	cmp	A complex cmp
F	Faraday constant	ΔH_{cmp}	Heat of formation of a complex cmp
R	Gas constant	mol _{cmp}	Number of moles of a complex cmp
T	Temperature (in Kelvin)	$Q_{(calc)}^c$	Calculated values of corrected heat from a calorimetric titration (see Eq. 12)
$E_{1/2S}$	Half wave potential of the metal ion in the absence of ligand	SSE	Sum of squared errors
$E_{1/2C}$	Half wave potential of the metal ion in the presence of ligand	Fitness	Variable describing how well a model encoded on a chromosome fits the experimental data (see Eq. 14)
I_{dS}	Diffusion limited current of the metal ion in the absence of ligand		
I_{dC}	Diffusion limited current of the metal ion in the presence of ligand		
$Fo_{(expt)}$	Fo term estimated from the <i>experimental</i> variables of a polarographic study (see Eq. 7)		
num	Number of measurements performed in the polarographic study		
y	A measurement y in a particular experiment		
$Q_{(expt)}$	Heat produced in a reaction vessel during a calorimetric titration		
C_{pr}	Heat capacity of the empty reaction vessel		

Acknowledgements

The authors would like to gratefully thank Albert Bos for many engaging and informative conversations. The authors would also like to thank the Royal Netherlands Government for providing an exchange grant for M.H. to work at the Twente University of Technology.

References

- [1] H.S. Rossotti, *Talanta*, 21 (1974) 809.
- [2] D.P.H. Zollinger, *Computerized Methods For The Determination Of Stability Constants*, Ph.D. Thesis, University of Twente, 1986.
- [3] A.L. Horvath, *Handbook of Aqueous Electrolyte Solutions: Physical Properties, Estimation and Correlation Methods*, Ellis Horwood, Chichester, 1985, p. 213.
- [4] G. Michaux and J. Reisse, *J. Am. Chem. Soc.*, 104 (1982) 6895.
- [5] C. Boudon, F. Peter and M. Gross, *J. Electroanal. Chem. Interfacial Electrochem.*, 117 (1981) 65.
- [6] H.S. Gold and M.R. Rice, *Talanta*, 29 (1982) 637.
- [7] P.U. Frueh, J.T. Clerc and W. Simon, *Helv. Chim. Acta*, 54 (1971) 1445.
- [8] F. Gaizer, *Coord. Chem. Rev.*, 27 (1979) 195.
- [9] F.J.C. Rossotti, H.S. Rossotti and R.J. Whewell, *J. Inorg. Nucl. Chem.*, 33 (1971) 2051.
- [10] I.G. Sayce, *Talanta*, 15 (1968) 1397.
- [11] R.M. Alcock, F.R. Hartley and D.E. Rogers, *J. Chem. Soc., Dalton Trans.*, (1978) 115.
- [12] C.B. Lucasius, *Towards Genetic Algorithm Methodology in Chemometrics*, Ph.D. Thesis, Katholieke Universiteit Nijmegen, 1993.
- [13] D.D. DeFord and D.N. Hume, *J. Am. Chem. Soc.*, 73 (1951), 5321.
- [14] D.J. Leggett, *Talanta*, 27 (1980), 787.
- [15] J.J. Christensen, J. Ruckman, D.J. Eatough and R.M. Izatt, *Thermochim. Acta*, 3 (1972), 203.
- [16] D.J. Eatough, J.J. Christensen and R.M. Izatt, *Thermochim. Acta*, 3 (1972) 219.
- [17] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
- [18] J.H. Holland, *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge, MA 1992.
- [19] A.P. De Weijer, C.B. Lucasius, L. Buydens, G. Kateman, H.M. Heuvel and H. Mannee, *Anal. Chem.*, 66 (1994), 23.
- [20] J.E. Baker, *Adaptive Selection Methods For Genetic Algorithms in J.J. Grefenstette (Ed.), Proceedings of the First International Conference on Genetic Algorithms*, Lawrence Erlbaum Assoc., Hillsdale, NJ, 1985, p. 101.
- [21] L.J. Eshelman and J.D. Schaffer, *Preventing Premature Convergence By Preventing Incest*, in R.K. Belew and L.B. Booker (Eds.), *Proceedings of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufman, San Mateo, CA, 1991, p. 115.
- [22] M.K. Hartnett, *The Application of Artificial Neural Networks and Genetic Algorithms To The Estimation of Electrode Response Characteristics and Stability Constants*, Ph.D. thesis, Dublin City University, 1994.
- [23] W. Edwards Deming, *Statistical Adjustment of Data*, Dover Publications, New York, 1964.
- [24] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in C (2nd edn.)*, Cambridge University Press, 1992; (a) *Numerical Differentiation* 186, (b) *Singular Value Decomposition* 59.
- [25] W.E. Wentworth, *J. Chem. Educ.*, 42 (1965) 96.
- [26] E. Casassus, R. Tauler and M. Filella, *Anal. Chim. Acta*, 191 (1986) 399.
- [27] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte and L. Kaufman, *Chemometrics: A Textbook*, Elsevier, Amsterdam, 1988.
- [28] M. Bos and H.Q.J. Meershoek, *Anal. Chim. Acta*, 61 (1972) 50.
- [29] G.A. Heath and G. Hefter, *J. Electroanal. Chem.*, 84 (1977) 295.
- [30] A. Laouenan and E. Suet, *Talanta*, 32 (1985) 245.
- [31] A.M. Bond, *Coord. Chem. Rev.*, 6 (1971) 377.
- [32] L. Chen, M. Bos, P.D.J. Grootenhuis, A. Christenhusz, E. Hoogendam, D.N. Reinhoudt and W.E. Van Der Linden, *Anal. Chim. Acta*, 201 (1987) 117.
- [33] D.J. Eatough, R.M. Izatt and J.J. Christensen, *Thermochim. Acta*, 3 (1972) 233.
- [34] M.A. Sharaf, D.L. Illman and B.R. Kowalski, *Chemometrics*, Wiley, New York, 1986.
- [35] K. Deb and D.E. Goldberg, *mGA in C: A Messy Genetic Algorithm in C*, IlliGAL Report No. 91008.
- [36] I. Harvey, *The SAGA Cross: The mechanics of recombination for species with variable length genotypes*, in R. Männer and B. Manderick (Eds.), *Parallel Problem Solving From Nature 2*, Elsevier, Amsterdam, 1992.
- [37] J.R. Koza, *Genetic Programming: A Paradigm for Genetically Breeding Populations of Computer programs to Solve Problems*, Stanford University Report STAN-CS-90-1314, 1990.
- [38] D.E. Goldberg and K. Deb, *A Comparative Analysis of Selection Schemes Used In Genetic Algorithms*, TCGA Report No. 90007.