Acta Chemica Scandinavica, 1996: 50: 717–734
Printed in UK – all rights reserved

Copyright © Acta Chemica Scandinavica 1996

ACTA
CHEMICA SCANDINAVICA
ISSN 0904–213X

LAKE – A Program System for Equilibrium Analytical Treatment of Multimethod Data, Especially Combined Potentiometric and Nuclear Magnetic Resonance Data

Nils Ingri,[a] Ingegärd Andersson,[a] Lage Pettersson,[a,*] Atsushi Yagasaki,[b] Lars Andersson[c] and Kenneth Holmström[d]

[a]Department of Inorganic Chemistry, Umeå University, S-901 87 Umeå, Sweden, [b]School of Science, Kwansei Gakuin University, 1-155 Uegahara Ichibancho, Nishinomiya 662, Hyogo, Japan, [c]Department of Mathematics, The Royal Institute of Technology, S-100 44 Stockholm, Sweden and [d]Department of Mathematics and Physics, Mälardalen University, P.O. Box 883, S-721 23 Västerås, Sweden

Ingri, N., Andersson, I., Pettersson, L., Yagasaki, A., Andersson, L. and Holmström, K., 1996. LAKE – A Program System for Equilibrium Analytical Treatment of Multimethod Data, Especially Combined Potentiometric and Nuclear Magnetic Resonance Data. – Acta Chem. Scand. 50: 717–734. © Acta Chemica Scandinavica 1996.

A computer program system called LAKE for equilibrium analytical treatment of aqueous multicomponent multiphase systems is described. The program package permits combined use of:

(1) potentiometric data: titration (EMF) and point (POT) data;
(2) nuclear magnetic resonance data: peak integral (NMRI) and chemical shift (NMRS) data;
(3) multiphase data: aqueous solution equilibrated with (i) a gas phase (GAS), (ii) solid phases (SOL) and (iii) surface phases (SURF);
(4) data at different pressures (P), temperatures (T) and ionic strengths (I);
(5) modelling data: for species modelling based on known chemical models;
(6) spectrophotometric data (SPEFO).

The package is portable and has been structured in a way which allows extension to further data types. The possibility of using multimethod data in LAKE enables the user to analyze systems of significantly higher complexity than is feasible with other program systems. Case studies illustrating the strength of multimethod analysis with the use of EMF, POT, NMRI and NMRS data for the treatment of a complicated polyanion system are discussed. The use of the package is illustrated by some short input and output tables. The core of optimization routines are based on new techniques for separable nonlinear least-squares problems.

1. Introduction

In the present paper the program package LAKE for equilibrium analytical treatment of systems with multimethod data is presented and discussed. The package at present permits combined treatment of the following types of data:

(1) Potentiometric: titration data (EMF) and point data (POT), i.e $-\log\{cx_i\}$;
(2) nuclear magnetic resonance: peak integral data (NMRI) and chemical shift data (NMRS);
(3) data with additional phases: i.e. systems where the aqueous phase is equilibrated with (i) species in a gas phase, (ii) solid phases and (iii) surface phases;
(4) data at different pressures (P), temperatures (T) and ionic strengths (I);
(5) spectrophotometric data (SPEFO).

Particular emphasis is placed on the possibilities of using a multilevel optimization technique for selecting 'best' equilibrium models in unknown aqueous multicomponent systems by using these types of data.

The first computer programs for solving equilibrium analytical problems were prepared by chemists and used mainly as a supplement to graphical methods. However, since increasingly complicated systems had to be treated, program systems of more general character were successively developed. In particular, different kinds of least-squares programs came into common use i.e. LETAGROP,[1] MINIQUAD,[2] SUPERQUAD,[3] HYPERQUAD (Ref. 4, pp. 402 and 403) and POLET

* To whom correspondence should be addressed.

(Ref. 5, pp. 145 and 146). Compilations of available least-squares programs for calculating formation constants can be found in Refs. 5 and 6. In Ref. 7 the different types of programs have been examined and discussed, in particular with reference to their mathematical–numerical functioning. In a recent book on data fitting,[8] one chapter deals with 'Potentiometric Titrations' and includes the historical development and an interesting discussion of how to obtain the best constants, using the computer program SUPERQUAD, in particular.

Of the different programs, LETAGROP, based on the so-called pit mapping method, was found to be useful especially in equilibrium systems with polynuclear species. These early computational developments were started in 1958 by Sillén and coworkers at The Royal Institute of Technology, KTH, in Stockholm.[9–11] Parallel to this development, algorithms for solving the mass balance equations also evolved. These were primarily aimed at computing normalized curves and projection maps, $Z(\log[H^+])_B$-curves and species distribution curves.[12]

The LETAGROP programs were primarily focused on least-squares determination of 'best' formation constants from various types of experimental data. Programs were developed for potentiometric titrations,[13] spectrophotometric measurements,[14] calorimetric titrations,[15] distribution equilibria,[16] kinetic data (not published) and for colligative properties.[17] In principle, each program treated solely one-method data, and no attempts were made to treat combined data simultaneously. The LETAGROP program package is still in use. However, the programs have been in use for almost 30 years and must now be considered obsolete.

The development of new experimental techniques has sharpened the equilibrium analysis considerably. NMR in combination with EMF has been found to be very powerful when evaluating complicated equilibrium systems. For instructive examples see Refs. 18 and 19. These considerations have motivated the development of a program system capable of unique evaluation of the 'best' equilibrium model from combined experimental data.

In 1984 a joint project was started in Umeå between the Department of Inorganic Chemistry and the Institute of Information Processing, aimed at developing a computational system for general treatment of complicated aqueous multiphase multicomponent systems with particular reference to the use of multimethod data. The new computational system, which is called LAKE, is presented below. A first description of the main ideas behind the computational unit was given in a paper by Andersson *et al.*[21] and the mathematical and numerical basis of the program by Andersson and Holmström.[22] Refs. 21 and 22 form parts of a thesis by Holmström.[7] DOS, PC/Windows and UNIX versions of the LAKE program package have been developed.

2. Basic equilibrium analytical concepts – applications in the LAKE system

All equilibrium analytical calculations are in principle based on the same fundamental concepts and formulae. These concepts form the general computational basis in the LAKE system, and are described below.

2.1. The equilibrium reaction. The equilibrium reaction states the type of chemical equilibria that will be used for describing a given chemical system. It gives the relationship between components A, B, C, ... and formed complexes $A_pB_qC_r$..., often denoted p,q,r,... . As in LETAGROP, the equilibrium reaction is defined as a formation reaction:

$$pA + qB + rC + ... \rightleftharpoons A_pB_qC_r... \qquad (I)$$

For clarity phase notations in reaction (I) have been omitted but will be commented on below. From now on, the general writing p,q,r, ... will be given in the abbreviated form pqr.

2.1.1. Use of the pqr-*concept.* Depending on how the components have been chosen, the pqr-values may attain positive or negative values, which means that the component in question is consumed or formed, respectively. Negative values are for instance found when component A represents protons and a deprotonization or hydrolysis reaction is considered. The components A, B, C in eqn. (I) are denoted (1,0,0), (0,1,0) and (0,0,1). A complex, for instance the decavanadate ion $V_{10}O_{28}^{6-}$ can, depending on how the components have been chosen, be written (−16,10,0) or (4,10,0) or (14,10,0). Additional illustrative examples are found in Table 1. The pqr-formalism has been shown to be a very convenient chemical basis for general mathematical–numerical treatment of complicated equilibrium systems.

2.1.2. Components, complexes and species. Following a commonly used nomenclature, both components and complexes will be denoted as species. As a rule, chemically well characterized, preferably mononuclear, species should be chosen as components. The number of components is defined as the smallest number of species by which all other species in the solution can be expressed, i.e. the components are the independent variables being used when solving the mass balance equations. When H^+ is chosen as component, OH^- is considered as a complex and vice versa. Even solid phases, gas molecules and sites on a surface will be referred to as species, and can thus be chosen as components. During an equilibrium analytical calculation it is sometimes advantageous, for chemical as well as numerical reasons, to make a new choice of components. Therefore, a routine has been worked out by which one can choose components arbitrarily among the species in an equilibrium model. Examples of some typical component choices are given in Table 1. This table illustrates that log β-values can

Table 1. Equilibrium model exemplifying species, *pqr*-notations and formation constants.[a]

Species			H^+–HVO_4^{2-}			H^+–$H_2VO_4^-$			H^+–VO_2^+			H^+–$V_{10}O_{28}^{6-}$		
Formula	Charge	Phase	*p*	*q*	log β	*p*	*q*	log β	*p*	*q*	log β	*p*	*q*	log β
H	+1	1	1	0	0.000	1	0	0.000	1	0	0.000	1.0	0.0	0.000
HVO_4	−2	1	0	1	0.000	−1	1	−7.920	−3	1	−14.880	−1.4	0.1	−13.133
H_2VO_4	−1	1	1	1	7.920	0	1	0.000	−2	1	−6.960	−0.4	0.1	−5.213
V_2O_7	−4	1	0	2	0.670	−2	2	−15.170	−6	2	−29.090	−2.8	0.2	−25.596
HV_2O_7	−3	1	1	2	10.590	−1	2	−5.250	−5	2	−19.170	−1.8	0.2	−15.676
$H_2V_2O_7$	−2	1	2	2	18.610	0	2	2.770	−4	2	−11.150	−0.8	0.2	−7.656
V_4O_{13}	−6	1	2	4	22.800	−2	4	−8.880	−10	4	−36.720	−3.6	0.4	−29.732
HV_4O_{13}	−5	1	3	4	31.900	−1	4	0.220	−9	4	−27.620	−2.6	0.4	−20.632
V_4O_{12}	−4	1	4	4	41.680	0	4	10.000	−8	4	−17.840	−1.6	0.4	−10.852
V_5O_{15}	−5	1	5	5	51.980	0	5	12.380	−10	5	−22.420	−2.0	0.5	−13.685
$V_{10}O_{28}$	−6	1	14	10	131.330	4	10	52.130	−16	10	−17.470	0.0	1.0	0.000
$HV_{10}O_{28}$	−5	1	15	10	137.330	5	10	58.130	−15	10	−11.470	1.0	1.0	6.000
$H_2V_{10}O_{28}$	−4	1	16	10	141.070	6	10	61.870	−14	10	−7.730	2.0	1.0	9.740
$H_3V_{10}O_{28}$	−3	1	17	10	142.700	7	10	63.500	−13	10	−6.100	3.0	1.0	11.370
VO_2	+1	1	3	1	14.880	2	1	6.960	0	1	0.000	1.6	0.1	1.747

[a] Four examples of component choices are given. In LAKE it is possible to arbitrarily choose components among the species in a given equilibrium model. Even solid phases and gas molecules can be used as components. Data are taken from a two-component study by Pettersson *et al.*[18]

change drastically at each change of components. Note that integers are not necessary to define a complex.

2.1.3. Phase conditions. The most common case is that all species are present in an aqueous phase (aq). In LAKE it is assumed that an aqueous phase is always present, but coupled equilibria between aqueous species, species in a gas phase (g), solid phases (s) and species on surfaces of solid phases (surface complexes) may also be treated.

2.2. The law of mass action and the formation constant. The concentration relations between components and formed complexes are determined by the law of mass action. Applying the law to equilibrium reaction (I) gives

$$[A_pB_qC_r\ ...] = \beta_{pqr} a^p b^q c^r\ ... \qquad \text{(II)}$$

where $[A_pB_qC_r\ ...]$ and $a,b,c,\ ...$ are the concentrations of complexes and components respectively, and β_{pqr} is the formation (stability) constant for the complexes formed. The constants are dependent on pressure (P), temperature (T) and the ionic strength (I), (in LAKE named as PTI-dependences). For cases with surface complexes or polyelectrolytes, β_{pqr} is also charge dependent.

2.2.1. Conditions at constant ionic medium. The β-constant defined according to eqn. (II) is based on an activity scale assuming quotients of activity coefficients to be constant and included in the β_{pqr}-values. This assumption is always used when applying the constant ionic medium method. This method implies that the equilibrium analysis is performed in an 'inert' salt medium of high total concentration. Commonly used media when studying anion equilibria are, for instance, 3.0 M $Na(ClO_4)$ and 0.6 M Na(Cl). In these media the Na^+ concentration is kept constant whereas the ClO_4^- and Cl^- ions can be partly replaced. In LAKE, as in LETAGROP, all mass balance summations are based upon relation (II), and thus on the fact that concentrations can be used instead of activities when applying the Law of Mass Action.

2.2.2. Treatment of PTI dependences. In LAKE it is possible to account for different pressure, temperature and ionic strength. The deviations from a given reference state are modelled by use of the equations defined in Table 2. These equations are based upon well-known thermodynamic relationships.

2.2.3. Treatment of Z, the charge dependence. For cases with surface complexation reactions according to the equilibrium

$$pH^+ + q(\equiv S{-}OH) + rC + ... \rightleftharpoons H_p(\equiv S{-}OH)_qC_r\ .... \qquad \text{(Ia)}$$

the law of maction gives

$$[H_p(\equiv S{-}OH)_qC_r\ ...] = \beta^{surf}_{pqr} h^p [\equiv S{-}OH]^q c^r\ .... \qquad \text{(IIb)}$$

where $[\equiv S{-}OH]$ is the current concentration of 'free' sites on the active surface, (defined as the component concentration), and $[H_p(\equiv S{-}OH)_qC_r\ ...]$ is the corresponding concentration of the formed surface complexes. The value of β^{surf}_{pqr} will vary depending on ΔZ_{pqr}, the change in charge of the surface caused by formation of the surface complex *pqr* and must be accounted for.

Charged groups of complexes on the surface of a solid phase give rise to a surface potential, Ψ. The coulombic energy, ΔG°_{coul}, required to bring one mole of ions from the bulk of solution to the surface at a potential Ψ, is given by

$$\Delta G^\circ_{coul} = \Delta Z_{pqr} F \Psi$$

This term has to be included in β^{surf}_{pqr} in order to correct

Table 2. PTI and *Z* (charge) dependences of the formation constants β_{pqr}.

Below are given the linear relationships, at present used in the LAKE system, for adjusting β_{pqr} due to *PTI* and *Z* (Charge) dependences. *P*=pressure, *T*=temperature, *I*=ionic strength and *Z*=charge of a surface active solid phase.
For a given solution, *m*, the log β-value for a species, *pqr*, is assumed to follow the relationship:

$\log \beta^{m}_{pqr} = \log \beta^{ref}_{pqr} + \Delta \log \beta^{m}_{pqr}$

$\log \beta^{ref}_{pqr}$ ... The log β-value at the reference state.

$\Delta \log \beta^{m}_{pqr}$ ... The modelling Δ-function used for describing the *PTI*-dependences of β^{ref}_{pqr}. The function is a sum of linear Δ-terms. The following four types of Δ-terms are at present available in the LAKE-system:

$\Delta \log \beta^{m}_{pqr} = \Delta(P, T)_m + \Delta(T)_m + \Delta(I, T)_m + \Delta(\Delta Z, T)_m$

The Δ-terms are the model functions describing *PTI* and charge (*Z*) dependences. Below are given the structure of the linear Δ-functions and explanations of the variables used. Note that in the table the point variables are denoted *m*.

(i) the Δ(PTI) model functions:

$$\Delta(P, T)_m = (\mathrm{PPAR})_1 \left[\frac{P}{T} - \frac{P_{ref}}{T_{ref}}\right]_m + (\mathrm{PPAR})_2 [T \log P - T_{ref} \log P_{ref}]_m + (\mathrm{PPAR})_3 [T^2 - T_{ref}^2]_m$$

$$\Delta(T)_m = (\mathrm{TPAR})_1 \left[\frac{1}{T} - \frac{1}{T_{ref}}\right]_m + (\mathrm{TPAR})_2 [\log T - \log T_{ref}]_m + (\mathrm{TPAR})_3 [T - T_{ref}]_m + (\mathrm{TPAR})_4 [T^2 - T_{ref}^2]_m$$

$$\Delta(I, T)_m = (\mathrm{IPAR})_1 \left[\frac{\sqrt{I}}{1 + a\sqrt{I}} - \frac{\sqrt{I_{ref}}}{1 + a\sqrt{I_{ref}}}\right]_m + (\mathrm{IPAR})_2 [I - I_{ref}]_m + (\mathrm{IPAR})_3 [I^{3/2} - I_{ref}^{3/2}]_m + (\mathrm{IPAR})_4 \left[\sqrt{I}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right) - \sqrt{I_{ref}}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right)\right]_m$$

PPAR TPAR IPAR ... Adjustable *P*, *T* and *I* parameters. The number of parameters NPPAR, NTPAR and NIPAR needed for describing the *PTI*-models are given in the LAKEDEF input file.

a ... Factor in the denominator of the Debye–Hückel term. Can be arbitrarily choosen and is not adjustable (common values are 1 or 1.5).

P_{ref} T_{ref} I_{ref} ... Reference state of *P*, *T* and *I*, respectively.

I ... An overall ionic strength in the *m*th solution, calculated using the expression $I = I_{med} + I_{pqr}$, i.e. the contributions from the medium ions and equilibrium species, respectively.

(ii) *The Δ(Z, T) model function* (the 'Helmholz Constant – Capacitance Model').[20] The model implies that the surface potential ψ is given by $\psi = \sigma\kappa^{-1}$.

κ ... The adjustable specific capacitance parameter (F m^{-2}). Named ZPAR in the program in analogy with PPAR, TPAR and IPAR.

σ ... The specific surface charge (C m^{-2}). Experimentally available through the relation:
$\sigma = \mathrm{SUMCHA} \cdot F(\mathrm{SOLIDC} \cdot \mathrm{SPAREA})^{-1}$

SUMCHA ... Surface charge (mol L^{-1} bulk solution), defined as $\sum z_{pqr}\, c^{sur}_{pqr}$

SOLIDC ... Concentration of solid active surface phase (g L^{-1}), setwise not adjustable parameter.

SPAREA ... Specific surface area (m^2 g^{-1}), a setwise adjustable parameter.

Combination of the formulae given in section 2.2.3 and the constant capacitance condition $\psi = \sigma\kappa^{-1}$, gives:

$$\Delta(\Delta Z, T) = -\frac{\Delta Z_{pqr}}{\kappa} \frac{F}{R \ln 10} \left(\frac{\sigma}{T} - \frac{\sigma_{ref}}{T_{ref}}\right)_m$$

ΔZ_{pqr} ... Change in charge of the surface species for the reaction.

σ_{ref} ... Refers to a reference state assuming an uncharged surface. The corresponding β^{surf}_{pqr} obtained is the 'intrinsic' constant. However, σ_{ref} and the corresponding $\log \beta^{int}_{pqr}$ can in the program be arbitrarily chosen.

for the charge dependence. This is achieved by splitting the overall observable Gibbs free energy, ΔG^{surf}_{pqr}, into ΔG°_{pqr}, an intrinsic free energy term and ΔG°_{coul}, a variable electrostatic or coulombic term. The following relationship is then valid:

$$\log \beta^{surf}_{pqr} = \log \beta^{int}_{pqr} - \Delta Z_{pqr} F \Psi / RT \ln 10 \qquad \text{(IIc)}$$

The Ψ-term can be considered as the charge dependence of the equilibrium constant β^{int}_{pqr}. By analogy with the β-dependence of *P*, *T* and *I*, given in Table 2, the general relationship

$$\log \beta^{surf}_{pqr} = \log \beta^{ref}_{pqr} + \Delta(-\Delta Z_{pqr}, \Psi, T)$$

is valid, where the Δ-term is the difference in log β from an arbitrarily chosen reference state log β^{ref}_{pqr}. The Δ-term may differ depending on the model used for describing the charge distribution from surface to bulk solution.

2.2.4. Summary of the combined treatment of PTI- *and Z-dependences of* β_{pqr}. The relationship between β_{pqr}, β^{ref}_{pqr} and their *PTI*- and *Z*-dependences can be written as

$$\log \beta_{pqr} = \log \beta^{ref}_{pqr} + \Delta(P,T) + \Delta(T) + \Delta(I,T) + \Delta(-\Delta Z,\Psi,T) \quad \text{(IId)}$$

where the Δ-terms are the linear functions given and explained in Table 2. It may be pointed out that β_{pqr} defined according to eqn. (IId), can be considered as a stoichiometric equilibrium constant and β^{ref}_{pqr} as a corresponding 'thermodynamic' equilibrium constant. Note that the 'best' β-constant in a LAKE refinement is β^{ref}_{pqr}, the extrapolated value as the Δ-terms→0.

2.3. The equilibrium model. The equilibrium model is defined as a species matrix including all species used for describing the speciation in the system studied. For each species, composition indices (*pqr*), formation constants (log β_{pqr}) and parameters needed for modelling the variations in pressure, temperature, ionic strength and charge are given. Codes for phase and isomer conditions (PHASE and ISOMER) are given as well. A detailed example of a model matrix is given in Table 3.

2.4. The mass balance equations (MBE). These equations give the conditions for the total concentrations of the components A, B, C, ... By summing all species in the equilibrium model the conditions for the total concentrations of A, B, C, ... are given by:

$$\begin{aligned} [A] &= a + \sum pcx_{pqr} = a + \sum p\beta_{pqr}a^p b^q c^r \ldots \\ [B] &= b + \sum qcx_{pqr} = b + \sum q\beta_{pqr}a^p b^q c^r \ldots \\ [C] &= c + \sum rcx_{pqr} = c + \sum r\beta_{pqr}a^p b^q c^r \ldots \\ &\vdots \end{aligned} \quad \text{(III)}$$

where the summation is taken over all complexes formed. cx_{pqr} and and *a*,*b*,*c*, ... are equilibrium concentrations of the complexes and the components, respectively.

2.4.1. Cases with surface complexation and solid phases. In systems with surface complexation and solid phases, MBE given in (III) must be complemented with additional terms. The total concentrations of the active sites on the surface, [B], is given by

$$[B] = \text{SURCON} = b + \sum q\beta^{surf}_{pqr}a^p b^q c^r \ldots \quad \text{(IIIa)}$$

and the summation is taken over all the surface complexes assumed to be formed. By analogy, the mass balances of *a* and *c* are then given by

$$\begin{aligned} [A] &= a + \sum pcx_{pqr} + \sum pcx^{surf}_{pqr} \\ [C] &= c + \sum rcx_{pqr} + \sum rcx^{surf}_{pqr} \end{aligned} \quad \text{(IIIb)}$$

The first sum in these expressions is taken over the aqueous complexes, and the second sum over the surface complexes.

In the case of solid phases, MBE must be complemented with cx^{solid}_{pqr}, i.e. the concentration of precipitated solid phases. Then for each equation

$$\sum pcx^{solid}_{pqr}, \sum qcx^{solid}_{pqr}, \sum rcx^{solid}_{pqr}, \ldots$$

has to be added.

The total concentrations as well as species concentrations are expressed in moles per litre of the bulk solution.

2.5. 'Best' formation constants. 'Best' formation constants are those that for a given speciation model give the best explanation of a given group of experimental data. As 'best' β_{pqr} are considered those minimizing a group of error squares sums U_1, U_2, ... The way of defining error squares sum is

$$U_k = \sum [w^k(y^k{}_{calc} - y^k{}_{obs})]^2 = \sum (w^k r^k)^2 \quad \text{(IV)}$$

where y_{calc} and y_{obs} are the calculated and observed quantities, respectively, r^k is the residual and *w* is a weighting factor for the points in the group *k*. The summation is taken over all points in the experimental group *k* considered. In LETAGROP only one error squares sum can be minimized. Moreover, the choice is restricted and as a rule H, the mass balance corresponding to the total proton concentration, is used. LAKE permits a totally general choice of residual sets. Residuals for the total concentrations (U_{MBE}) as well as residuals for the different measured equilibrium quantities (U_{MEQ}) may be used. This implies minimization of

$$U_{sum} = \sum_{i=1}^{NCOM} U_{MBE}(i) + \sum_{i=1}^{NMEQ} U_{MEQ}(i) \quad \text{(IVa)}$$

where NCOM and NMEAS are number of mass balance equations and number of measured equilibrium quantities, respectively. These residual possibilities make the LAKE system particularly useful for selective search in unknown equilibrium systems. Moreover, the choice of residuals and weightings makes it possible to use all available information in an optimal way and also to control the influences of different experimental parameters.

2.5.1. On the minimization method. In LAKE the main optimization algorithm is a separable Gauss–Newton method,[23] but it is the non-linear parameters (e.g. free concentrations) that are separated out, as opposed to the most often discussed case where the linear parameters (e.g. formation constants, shift parameters) are separated out. LAKE uses general purpose numerical software. LINPACK[24] is used for linear algebra computations and a modified version of LSUNC[25] is used (in two nearly identical versions) to implement the separable Gauss–Newton algorithm. The step length algorithm in LSUNC is changed to a modified version of an algorithm by Fletcher.[26] All derivatives are calculated analytically.

The separable Gauss–Newton algorithm is similar in

Table 3. A numerical MODDEF input example.
The input example illustrates the MODDEF input of an equilibrium model with 4 components and 47 complexes. The input is based on the following type of equilibria $pH^+ + qMoO_4^{2-} + rHVO_4^{2-} + sHPO_4^{2-} \rightleftharpoons (H^+)_p(MoO_4^{2-})_q(HVO_4^{2-})_r(HPO_4^{2-})_s$

LGBETA	FCPHAS	ISOMER	*p*	*q*	*r*	*s*	Comments
0.00	1	1	1.00	0.00	0.00	0.00	H^+
0.00	1	1	0.00	1.00	0.00	0.00	MoO_4^{2-}
0.00	1	1	0.00	0.00	1.00	0.00	HVO_4^{2-}
0.00	1	1	0.00	0.00	0.00	1.00	HPO_4^{2-}
3.39	1	1	1.00	1.00	0.00	0.00	$H_p(MoO_4)_q$
7.34	1	1	2.00	1.00	0.00	0.00	
52.42	1	1	8.00	7.00	0.00	0.00	*pq*00
57.22	1	1	9.00	7.00	0.00	0.00	
60.78	1	1	10.00	7.00	0.00	0.00	
71.63	1	1	12.00	8.00	0.00	0.00	
73.37	1	1	13.00	8.00	0.00	0.00	
76.34	1	1	15.00	8.00	0.00	0.00	
41.68	1	1	4.00	0.00	4.00	0.00	$H_p(HVO_4)_r$
131.33	1	1	14.00	0.00	10.00	0.00	
137.33	1	1	15.00	0.00	10.00	0.00	*p*0*r*0
7.92	1	1	1.00	0.00	1.00	0.00	
18.60	1	1	2.00	0.00	2.00	0.00	
14.88	1	1	3.00	0.00	1.00	0.00	
6.418	1	1	1.00	0.00	0.00	1.00	$H_p(HPO_4)_s$
8.190	1	1	2.00	0.00	0.00	1.00	*p*00*s*
134.56	1	1	15.00	1.00	9.00	0.00	$H_p(MoO_4)_q(HVO_4)_r$
137.33	1	1	16.00	1.00	9.00	0.00	
60.37	1	1	8.00	4.00	2.00	0.00	*pqr*0
64.11	1	1	9.00	4.00	2.00	0.00	
59.14	1	1	9.00	5.00	1.00	0.00	
74.63	1	1	11.00	7.00	1.00	0.00	
105.22	1	1	13.00	4.00	5.00	0.00	
107.58	1	1	14.00	4.00	5.00	0.00	
57.38	1	1	8.00	5.00	0.00	2.00	$H_p(MoO_4)_q(HPO_4)_s$
63.11	1	1	9.00	5.00	0.00	2.00	
67.49	1	1	10.00	5.00	0.00	2.00	*pq*0*s*
91.96	1	1	14.00	9.00	0.00	1.00	
96.07	1	1	15.00	9.00	0.00	1.00	
99.05	1	1	16.00	9.00	0.00	1.00	
100.98	1	1	17.00	9.00	0.00	1.00	
99.18	1	2	16.00	9.00	0.00	1.00	
101.63	1	2	17.00	9.00	0.00	1.00	
115.66	1	1	18.00	11.00	0.00	1.00	
118.93	1	1	19.00	11.00	0.00	1.00	
120.42	1	1	20.00	11.00	0.00	1.00	
131.67	1	1	23.00	12.00	0.00	1.00	
93.13	1	1	14.00	10.00	0.00	1.00	
98.23	1	1	15.00	10.00	0.00	1.00	
110.78	1	1	17.00	11.00	0.00	1.00	
141.48	1	1	22.00	11.00	1.00	1.00	$H_p(MoO_4)_q(HVO_4)_r(HPO_4)_s$
146.93	1	1	21.00	10.00	2.00	1.00	
147.19	1	1	22.00	10.00	2.00	1.00	*pqrs*
147.09	1	1	20.00	9.00	3.00	1.00	
150.19	1	1	21.00	9.00	3.00	1.00	
151.73	1	1	22.00	9.00	3.00	1.00	
152.22	1	1	23.00	9.00	3.00	1.00	

(i) The model input data shall always start with the components.
(ii) H^+ (or OH^-) shall always be given as the first component. When H^+ is component, OH^- is then considered as a complex and denoted $-1,0,0,0$, The formation constant ($\beta_{-1,0,0,0}$) is the ionic product of water.
(iii) For the components $\log \beta = 0$.
(iv) The ordering between the complexes can be chosen arbitrarily. It may, however, be convenient to arrange them according to their component content e.g. one-component species, two-component species etc.

spirit to that of SUPERQUAD,[3] the main difference being the way in which the constraints $\beta \geqslant 0$ on the formation constants are handled. In LAKE $\log \beta$ is optimized, which ensures positivity, in contrast to MINIQUAD[2] and SUPERQUAD.[3] LAKE also uses a numerically more stable algorithm. Bounds are set on all optimized parameters, which function as a regularization procedure.

2.6. The 'Best' model. The 'best' model is the equilibrium model that explains all parts of the concentration space and all sets of measured equilibrium concentrations. An explanation of a given group of data is defined as the 'best' set of model parameters fulfilling given requirements of the statistical errors. Species with $\sigma(\beta_{pqr}^{\mathrm{ref}}) \geqslant 0.66\ \beta_{pqr}$ can, for instance, be rejected (see the species selector discussion in Ref. 11). A common rule for equilibrium chemists has been to propose as 'best' model the model containing the smallest number of species. This is by no means a satisfactory rule. It seems more logical to use some species-related criteria for accepting and rejecting species. An important part in this species selection is the choice of data. Only species fulfilling suitable statistical criteria, such as the Akaike's Information Theoretic Criterion (AIC)[27] should be included in the equilibrium model. In applying these methods data must be divided into subsets and then the 'best' equilibrium model is determined for each of them. In a final step each of these submodels are then combined, compared and statistically evaluated. Through such a procedure it is possible to judge the quality of different data and the accuracy of the results obtained. In cases with multimethod data such an evaluation is particularly valuable.

3. Equilibrium measurements

In equilibrium analytical work it is a necessity that the experimental methods used have high selectivity, and can accurately measure equilibrium concentrations quantitatively over a broad range of concentrations. However, even with these requirements fulfilled, experience has shown that data from a single experimental method often is not enough for a unique explanation of a complicated multicomponent equilibrium system. For a successful analysis it is often necessary to have complementary data from two or more experimental methods.

3.1. Methods – selectivity and types of experimental data. The selectivity of a method is determined by the physico-chemical property upon which the method is based. Examples of such properties are light absorption (spectrophotometry) and heat evolution (calorimetry). The selectivity of the method is dependent on the kind of measured quantity that is obtainable, and on the function determining the relationship between the experimentally observed quantity, y_{obs}, and the corresponding set of species concentrations, $[cx_i]$. In the following discussion this function will be denoted as the measuring function.

For a general computer treatment of cases with multimethod data it is necessary to have a program system permitting combined treatment of many types of experimental data. In the LAKE system it was found convenient to group the different experimental methods into four main groups depending on the selectivity, physical properties and types of measuring functions. From a methodological point of view the four groups are in principle based on (1) the chemical potential, (2) subsets of total concentrations, (3) spectroscopic behaviour and (4) colligative properties. The four groups form the basis for the treatment of multimethod data in the LAKE system. A compilation of these four measuring situations is given in Table 4. The relationships between experimentally measured quantities, y_{obs}, and the corresponding species, cx_i, are given as well. As can be seen, most measuring situations are covered by those given for EMF, POT, NMRI/NMRS and SPEFO. In the present version of LAKE the first three, EMF, POT and NMRI/NMRS, have been completely worked out and are in continuous use since 1989. The SURF, SOLID, GAS and SPEFO types are newly developed. Data of KALLE and CONDU type should be analyzed using the SPEFO type, and data of OSMO type using the NMRI type. See Table 4 and the discussion in section 7.3.

4. The input structure of the LAKE program

The initial Input consists of System Description Data, Model Description Data and Experimental Multimethod Data. In the following these data types will be briefly discussed and the presentation kept very close to the vocabulary used in the program. Our intention is merely to point to the many possibilities with the program rather than to present a guide for the user.

4.1. The system description data – the LAKEDEF input file. Each new Chemical System is started by running the program LAKEDEF, which creates the base files used by LAKE. The input contains a Global Description concerning phase conditions, component choices, medium conditions and conditions for pressure, temperature and ionic strength. It also gives some special experimental information of, for instance, the NMR measurements and the presence of dirt acids.

4.2. The model description data – the MODDEF input file. The Model Description Data Input defines the initial equilibrium model used by the system. The input is a listing of all the species (i.e. components and complexes) included in the model. 'A Numerical MODDEF Input Example' is given in Table 3. The example illustrates an aqueous four-component model with many species, some isomeric species included. The molybdovanadophosphate system used to exemplify the data input is still under study. Some preliminary results are, however, reported

Table 4. Survey of equilibrium analytical measuring methods.

Measured *cx(i)* species	Measuring function	LAKE name	Observed quantities
Single species *cx(i)*	$y_{(i)}^{\mathrm{EMF}} = E_{(i)} = E_{(i)}^{0} + k_{\mathrm{EL}(i)}\, gn_{e}^{-1} \log\{cx(i)\}$	EMF (MSTYPE = 1)	Electrode potentials, $E_{(i)}$ $y_{\mathrm{obs}}^{(i)} = E_{(1)}, E_{(2)}, E_{(3)}, \ldots$
	$y_{(i)}^{\mathrm{POT(1)}} = \log\{cx(i)\}$	POT(1) (MSTYPE = 2)	Activities $\{cx(i)\}$, directly measured given as $\log\{cx(i)\}$ $y_{\mathrm{obs}}^{(i)} = \log\{cx(1)\}, \log\{cx(2)\}, \log\{cx(3)\}, \ldots$
	$y_{(i)}^{\mathrm{POT(2)}} = \log\{Y_{(i)}\} = Y_{(i)}^{0} + K_{RT(i)} \log\{cx(i)\}$	POT(2) (MSTYPE = 3)	Activities $\{cx(i)\}$, indirectly measured given as $\log\{Y_{(i)}\}$ $y_{\mathrm{obs}}^{(i)} = \log Y_{(1)}, \log Y_{(2)}, \log Y_{(3)}, \ldots$
Subsets of B_{tot}	$I_{\mathrm{NMRI}(j)}^{\mathrm{NMRI}} = B_{\mathrm{PEAK}(j)}^{\mathrm{NMRI}} = \mathrm{f}_{\mathrm{eq}_{(j)}}\, Q_{PQR(j)} \sum_{i=0}^{\mathrm{NSPEC}} cx(i)$	NMRI	NMR peak integrals, $I_{\mathrm{PEAK}(j)}^{\mathrm{NMRI}}$ $y_{\mathrm{obs}}^{(j)} = I_{\mathrm{PEAK}(1)}^{\mathrm{NMRI}}, I_{\mathrm{PEAK}(2)}^{\mathrm{NMRI}}, I_{\mathrm{PEAK}(3)}^{\mathrm{NMRI}}, \ldots$
	$D_{\mathrm{DIST}(j)}^{\mathrm{LIQ}} = B_{\mathrm{DIST}(j)}^{\mathrm{ORG}} = F_{\mathrm{VOL}(j)} \sum_{i=1}^{\mathrm{NSPEC}} q_{(i)}\, cx(i)$	DISTR	Liquid phase distributions, $D_{\mathrm{DIST}(j)}^{\mathrm{LIQ}}$ $y_{\mathrm{obs}}^{(j)} = D_{\mathrm{DIST}(1)}^{\mathrm{LIQ}}, D_{\mathrm{DIST}(2)}^{\mathrm{LIQ}}, D_{\mathrm{DIST}(3)}^{\mathrm{LIQ}}, \ldots$
	$D_{\mathrm{DIST}(j)}^{\mathrm{SURF}} = B_{\mathrm{DIST}(j)}^{\mathrm{SURF}} = \sum_{i=1}^{\mathrm{NSPEC}} q_{(i)}\, cx(i)$	SURF	Surface phase distributions, $D_{\mathrm{DIST}(j)}^{\mathrm{SURF}}$ $y_{\mathrm{obs}}^{(j)} = D_{\mathrm{DIST}(1)}^{\mathrm{SURF}}, D_{\mathrm{DIST}(2)}^{\mathrm{SURF}}, D_{\mathrm{DIST}(3)}^{\mathrm{SURF}}, \ldots$
	$S_{\mathrm{SOL}(j)}^{\mathrm{SOLID}} = B_{\mathrm{SOL}(j)}^{\mathrm{SOLID}} = \sum_{i=1}^{\mathrm{NSPEC}} q_{(i)}\, cx(i)$	SOLID	Solid phase solubilities, $S_{\mathrm{SOL}(j)}^{\mathrm{SOLID}}$ $y_{\mathrm{obs}}^{(j)} = S_{\mathrm{SOL}(1)}^{\mathrm{SOLID}}, S_{\mathrm{SOL}(2)}^{\mathrm{SOLID}}, S_{\mathrm{SOL}(3)}^{\mathrm{SOLID}}, \ldots$
	$S_{\mathrm{SOL}(j)}^{\mathrm{GAS}} = B_{\mathrm{SOL}(j)}^{\mathrm{GAS}} = \sum_{i=1}^{\mathrm{NSPEC}} q_{(i)}\, cx(i)$	GAS	Gas phase solubilities, $G_{\mathrm{SOL}(j)}^{\mathrm{GAS}}$ $y_{\mathrm{obs}}^{(j)} = S_{\mathrm{SOL}(1)}^{\mathrm{GAS}}, S_{\mathrm{SOL}(2)}^{\mathrm{GAS}}, S_{\mathrm{SOL}(3)}^{\mathrm{GAS}}, \ldots$
Linear combination of *cx(i)*	$\Delta_{\mathrm{PEAK}(j)}^{\mathrm{NMRS}} = \sum_{i=0}^{\mathrm{NSPEC}} \delta_{cx(i)}\, cx(i) \Big/ \sum_{i=0}^{\mathrm{NSPEC}} cx(i)$	NMRS	NMR chemical shift values, $\Delta_{\mathrm{PEAK}(j)}^{\mathrm{NMRS}}$ $y_{\mathrm{obs}}^{(j)} = \Delta_{\mathrm{PEAK}(1)}^{\mathrm{NMRS}}, \Delta_{\mathrm{PEAK}(2)}^{\mathrm{NMRS}}, \Delta_{\mathrm{PEAK}(3)}^{\mathrm{NMRS}}, \ldots$
	$\mathrm{OD}_{\mathrm{l}(j)}^{\mathrm{SPEFO}} = \sum_{i=1}^{\mathrm{NSPEC}} \varepsilon_{\mathrm{cx(i)}}^{\lambda}\, cx(i)$	SPEFO	Optical densities, $\mathrm{OD}_{\lambda(j)}^{\mathrm{SPEFO}}$ $y_{\mathrm{obs}}^{(j)} = \mathrm{OD}_{\mathrm{l}(1)}^{\mathrm{SPEFO}}, \mathrm{OD}_{\mathrm{l}(2)}^{\mathrm{SPEFO}}, \mathrm{OD}_{\mathrm{l}(3)}^{\mathrm{SPEFO}}, \ldots$
	$\mathrm{Q}_{\mathrm{cal}(j)}^{\mathrm{KALLE}} = \sum_{i=1}^{\mathrm{NSPEC}} \Delta H_{cx(i)}\, cx(i)$	KALLE	Reaction heats, $\mathrm{Q}_{\mathrm{cal}(j)}^{\mathrm{KALLE}}$ $y_{\mathrm{obs}}^{(j)} = \mathrm{Q}_{\mathrm{cal}(1)}^{\mathrm{KALLE}}, \mathrm{Q}_{\mathrm{cal}(2)}^{\mathrm{KALLE}}, \mathrm{Q}_{\mathrm{cal}(3)}^{\mathrm{KALLE}}, \ldots$
	$\Lambda_{(j)}^{\mathrm{CONDU}} = \sum_{i=1}^{\mathrm{NSPEC}} \lambda_{\mathrm{cx(i)}}\, cx(i)$	CONDU	Conductivities, $\Lambda_{(j)}^{\mathrm{CONDU}}$ $y_{\mathrm{obs}}^{(j)} = \Lambda_{(1)}^{\mathrm{CONDU}}, \Lambda_{(2)}^{\mathrm{CONDU}}, \Lambda_{(3)}^{\mathrm{CONDU}}, \ldots$
Osmotic sum of *cx(i)*	$\Delta_{(j)}^{\mathrm{OSMO}} = K^{\mathrm{OSMO}}\, S = K^{\mathrm{OSMO}} \sum_{i=1}^{\mathrm{NSPEC}} cx(i)$	OSMO	Vapour phase osmometry, $\Delta_{(j)}^{\mathrm{OSMO}}$ $y_{\mathrm{obs}}^{(j)} = \Delta_{(1)}^{\mathrm{OMSO}}, \Delta_{(2)}^{\mathrm{OMSO}}, \Delta_{(3)}^{\mathrm{OMSO}}, \ldots$

in Ref. 28. At present the program can treat up to 12 components and 120 complexes, and these numbers can easily be expanded.

4.3. The experimental multimethod data – data input file. The Multimethod Data File in LAKE is built up of consequtive sets (SET1, ...) and is read into the data base by the program NEWSET. The order of the sets can be chosen arbitrarily. However, it may be convenient to arrange the sets in groups of similar experimental type.

4.3.1. Description of the set structure. A set is built up of (i) Set Description Data, (ii) Basic Set Data, either EMF or POT, and if POT then (iii) Additional Data Types (NMRI, NMRS, SPEFO) can be added in one and the same set.

(i) Set Description Data. A new set always starts with a Set Description, which states the general experimental conditions valid for all data types used in the set including: number of points in the set (NPKT); statements concerning PTI-conditions; number of measured equilibrium concentrations (NMEAS); and for each MEAS a species description.

(ii) Basic Set Data. After the Set Description Data one gives the Basic Data Type. At present it is possible to use the two Basic Data Types EMF and POT, and only one of these two Basic Data Types can be used in one and the same set.

EMF refers to Potentiometric Titration Data, which implies that the total component concentrations are given by volume-additions in a titration procedure. After each addition equilibrium concentrations are measured by using the values of electrode potentials, $Y_{obs}=E$ (MSTYPE=1). In addition, it is also possible in an EMF set to have directly and/or indirectly measured equilibrium concentrations (MSTYPE=2 and 3).

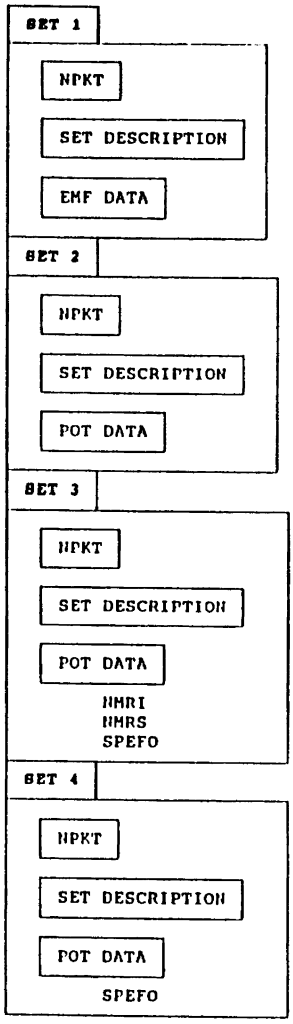
POT refers to Potentiometric Point Data with an Input consisting of total component concentrations and directly and/or indirectly measured equilibrium concentrations (MSTYPE=2 and 3).

(iii) Additional Data Types. Besides EMF and POT one can also use Additional Data Types and at present the following data types are defined:

NMRI, NMR peak integral data, NMRS, NMR chemical shift data and SPEFO, spectrophotometric absorbance data.

A flowsheet of a multimethod data file illustrating how the different data types can be mixed to form different types of sets is given in Table 5. In remarks (1)–(3) some general rules for constructing sets in a data file are given. Note that the restrictions in combining data are not caused by limitations in the program but are chemically motivated.

Table 5. Flowsheet exemplifying a multimethod data file.



(1) EMF DATA and POT DATA cannot be given in one and the same SET.
(2) NPKT (number of points) and SET DESCRIPTION must be given for each SET.
(3) NMRI, NMRS and SPEFO must always be combined with POT DATA, but cannot be combined with EMF DATA.

5. Potentiometric titration data – EMF

A potentiometric titration experiment implies that V_T ml of a solution in a burette is added to V_0 ml of a solution in a titration vessel. After each addition, electrode potentials (E_1, E_2, E_3, ...) are measured. The total component concentrations in the titration vessel (A_0, B_0, C_0, ...) and in the burette (A_T, B_T, C_T, ...) are assumed to be known. Then from V_0 and added values of V_T the total component concentrations after mixing (A, B, C, ...) can be calculated. Thus a titration experiment gives equilibrium data of the general type $(A, B, C, ..., E_1, E_2, E_3, ...)_{NPKT}$ where NPKT is the number of points in the titration.

The general feature of the algorithm used in LAKE for treating potentiometric titration data is in principle the same as in LETAGROP. While in LETAGROP the treatment is restricted solely to cases with aqueous solution, in LAKE it is now possible also to treat systems with Additional Phases. At present, titration experiments for cases where the aqueous phase is equilibrated with

(i) species in a gas phase (GAS), (ii) solid phases (SOLID) and (iii) surface phases (SURF) can be treated.

Besides the measured E-values in a titration procedure, it is also possible to have additional types of concentration measurements. At present directly measured activities given as $Y_{obs} = -\log\{\ \}$, and indirectly measured activities as $Y_{obs} = Y_0 + K_{RT}\log\{\ \}$ are handled.

5.1. The volume and concentration variables. Besides the basic concentration variables described above, additional variables have been introduced in the LAKE program: total concentrations of cations and anions in the medium, (used to calculate I_{med}, the ionic strength contribution due to the medium ions), and concentration contribution due to equilibrated additional phases, (i.e. components in a gas phase, solid phases and active surface phases).

These basic titration variables are considered as experimental quantities and are not adjustable during a computation. However, in order to make error estimates of the experimentally determined concentration variables possible, the following adjustable error variables are used: analytical errors of component concentrations in vessel and burette solutions, dirt acid errors (concentrations of proton dependent impurities initially present in burette and vessel solutions), and relative errors of total concentrations after mixing.

5.2. The EMF measurement and measuring variables. The basic measuring function is the Nernst equation, in principle used in the same way as in LETAGROP. Within an EMF-set two additional types of single species measurements (direct and indirectly measured activities) may be used as well, either alone or in combination with measured E-values. The relationship between the potentiometrically measured EMF value, E, and the corresponding equilibrium concentration, $[cx(i)]$, is given by the expression

$$E_{(i)} = E^{\circ}_{(i)} + \frac{g}{n_e}\log[cx(i)] + E_j$$

where $E^{\circ}_{(i)}$ is an experimental constant determined by the cell reaction, and g is a temperature-dependent coefficient which can be expressed as

$$g = \frac{R}{F}\ln 10\ T = 1.9842 \times 10^{-4}\,\mathrm{T}$$

where T is the absolute temperature, and R and F are the gas and the Faraday constants, respectively. n_e is the number of electrons taking part in the cell reaction, and E_j is the liquid junction potential. E_j is defined as $E_j = \Sigma\, k_{pqr}[cx_{pqr}]$, where the summation is taken over all pqr-species assumed to contribute to the liquid junction potential. With this definition it is possible to consider E_j-contributions not only from H^+ and OH^-, but also from additional ions present in the solution. In LAKE there are no restrictions on the number of electrodes that can be used in a titration. The EMF-measured species are always chosen as components. In order to correct for non-ideal behaviour of an electrode it is possible to multiply g with a multiplicative factor k_{EL}, adjustable and normally given as unity. The adjustable parameters E°, k_{EL} and E_j are electrode-related parameters and must be given for each electrode and in all EMF sets of the data file. These values can be obtained either through separate determinations, or by direct determinations using set data given either as a separate set or as a part of a set. The g-coefficient is directly calculated in the program by using the formula given above.

6. Potentiometric point data – POT

POT data are point data and differ in this respect from EMF data (titrations). POT data can be used either alone or in combination with additional data types, see the Flowsheet given in Table 5. POT sets are also used for modelling calculations.

6.1. Description of the POT data input. A POT set always starts with a Set Description followed by the POT data input, which consist of

(A) Number of Additional Phases: equilibrated gas-phase components (NGAS), solid phases (NSOLID), surface phases (NSURF).
(B) Measuring Parameters: indirectly measured activities and in case of surface phases specific areas and grams of added active solid phase.
(C) Raw Data. The input includes:
  (i) **total concentrations** of components, medium ions, dirt acids, surface sites (in moles L^{-1}), added solid phases (in moles L^{-1}) and added solid surface phases (in g L^{-1})
  (ii) **equilibrium concentrations** of components measured according to MSTYPE=2 and 3 given as $-\log\{\ \}$. When Additional Phases are present: the measured solubilities of equilibrated gas species, which need not be components (NGAS>0), and the equilibrium solubilities of initially added solid phases (NSOLID>0).

7. NMR point data – NMRI and NMRS

During the development of the NMR part we have made use of the experimental data of type EMF (potentiometric titrations), POT (pH point data), NMRS (shift data), NMRI (peak integral data) collected by Atsushi Yagasaki on the molybdophenylphosphonate system, H^+–MoO_4^{2-}–$C_6H_5PO_3^{2-}$, during his stay at the Department of Inorganic Chemistry, Umeå University. Data are presented in his thesis 1986,[29] and the final result of the study was later published.[30]

Besides the molybdophenylphosphonate system, data and results from e.g. the molybdovanadate[31] and the molybdophosphate[32] systems studied by Pettersson *et al.* have been extensively used for developing the NMR routines. As examples of measurements with two NMR-active nuclei we have also used data kindly provided by Julius Glaser, KTH, Stockholm. In this case the com-

plexation between Tl^+ and CN^- was studied in strongly acid solutions.[33]

In LAKE all kinds of equilibrium measurements are related to a global species model including all species in the equilibrium model. An example of such a model is given in Table 3. Subsets of the global model, in the NMR case submodels including only the NMR active species, can be defined as well. At present, it is possible to have one submodel for integral data (NMRI) and one for shift data (NMRS).

7.1. Peak integral data – NMRI. The aims of using peak integrals in equilibrium analysis have been to (i) increase the sharpness of the evaluation, (ii) find structurally different 'core' units in polynuclear species, (iii) determine isomeric species and (iv) determine medium dependence of formation constants.

7.1.1. Peak equilibria. Consider a set of NMR-active species of the following general composition, $H_PB_QL_R(H)_i(L)_j$, where B is the component assumed to contain the NMR-active nucleus, (e.g. ^{51}V or ^{31}P). $H_PB_QL_R$ is a core unit in common for all the (i,j)-species formed. The 'free' core species ($i=0, j=0$) is assumed to be formed by slow exchange reactions of the protons and the ligands. $(H)_i$ and $(L)_j$ are fast exchangeable protons and/or ligands coordinated to the core unit. In the vanadate system, for instance, the core unit $V_{10}O_{28}$ is present in all the decavanadate species formed $V_{10}O_{28}^{6-}$, $V_{10}O_{28}(H)^{5-}$, $V_{10}O_{28}(H)_2^{4-}$ and $V_{10}O_{28}(H)_3^{3-}$. The following four types of peak equilibria can be defined:

(0,0) Core formation (slow exchange)

$$PH + QB + RL \rightleftharpoons H_PB_QL_R \qquad (V)$$

$(i,0)$ Core protonization (fast exchange)

$$H_PB_QL_R + iH \rightleftharpoons H_PB_QL_R(H)_i \qquad (Va)$$

$(0,j)$ Core ligation (fast exchange)

$$H_PB_QL_R + jL \rightleftharpoons H_PB_QL_R(L)_j \qquad (Vb)$$

(i,j) Mixed proton-ligand formation (fast exchange)

$$H_PB_QL_R + iH + jL \rightleftharpoons H_PB_QL_R(H)_i(L)_j \qquad (Vc)$$

From this equilibrium description it is evident that the peak speciation is quantitatively determined by a formation constant β_{PQR} and a set of step constants k_i, k_j or $k_{i,j}$.

7.1.2. Peak mass balances. Since the core unit is homonuclear with respect to the NMR-active B-component and present in all species, the condition for total B gives

$$B_{PQR} = Q \sum\sum cx_{i,j} = Q\beta_{PQR}\, h^P b^Q c^R \sum\sum k_{i,j} h^i l^j$$

where Q is the number of B-atoms in the core unit $H_PB_QL_R$ ($i=0, j=0$) and $cx_{i,j}$ the species concentrations defined according to the general equilibria (Vc). Note that this general way of writing includes all peak species. For instance $cx_{0,0}$ corresponds to the concentration of the core species ($k_{0,0}=1$).

The relationship between the formation constant β_{pqr} and k_{ij} is given by the expression

$$k_{i,j} = \beta_{pqr}/\beta_{PQR}$$

where β_{pqr} is the overall formation constant, i.e. $p=P+i$, $q=Q$ and $r=R+j$.

7.1.3. Peak speciation. If we for instance consider general peak equilibria according to eqn. (Vc), a peak integral in an NMR spectrum always consists of a sum of B-concentrations. This sum is dependent on whether the B-atoms in the core unit $H_PB_QL_R$ are structurally equivalent or not. For the case where all the B atoms are structurally equivalent, the NMR spectrum contains a single peak and the peak integral is proportional to B_{PQR} and can be written

$$I_{PQR} = K_{sc}B_{PQR} = K_{sc}Q \sum\sum cx_{i,j}$$

where K_{sc} is an experimental scale factor determining the relationship between peak integrals and the summed concentrations of the species in the peak. For the case where the B-atoms are structurally different in the core unit, B_{PQR} is distributed over a number of peaks with distribution factors f_{eq}. The number of peaks will be equal to the number of structurally equivalent groups of B-atoms.

7.1.4. Peak integral example. In Fig. 1 the NMR-spectrum of the decavanadate unit $V_{10}O_{28}$ is shown. As seen, the decavanadates give rise to three peaks denoted V_1, V_2 and V_3. The ratios between the peak-integrals $V_1:(V_2+V_3)=1:4$ and $V_2:V_3=1:1$ giving a $V_1:V_2:V_3$ ratio of 1:2:2. This corresponds to the distribution factors (f_{eq}) of 0.2, 0.4 and 0.4. Thus, the NMR spectrum shows that the Q-value must be a multiple of 5. In calculations by combined use of peak integrals, pH-values and total concentrations, H and B, it is easy to establish $Q=10$. These results agree very well with the crystal structure of the decavanadate ion shown schematically as a part of Fig. 1.

7.2. Chemical shift data – NMRS. In the LAKE system it is possible to combine peak integral data (which give subsets of equilibrium concentrations) with chemical shift data. The main purpose is hereby to identify and determine proton and/or ligand series. From an equilibrium analytical point of view, the aim is to use and analyse shift profile data, i.e. functions $\delta(pH)$, where δ is the averaging of the individual shift values $\delta_{i,j}$.

7.2.1. The shift profile. In cases with solely core formation according to eqn. (V) the peak has a constant shift value and is independent of proton and/or ligand concentrations. In such cases a shift value does not give any equilibrium analytical information and can be used solely as a fingerprint of structurally equivalent B-atoms (δ_{PQR}=constant). However, for cases where the core formation reaction is coupled also to fast proton and/or

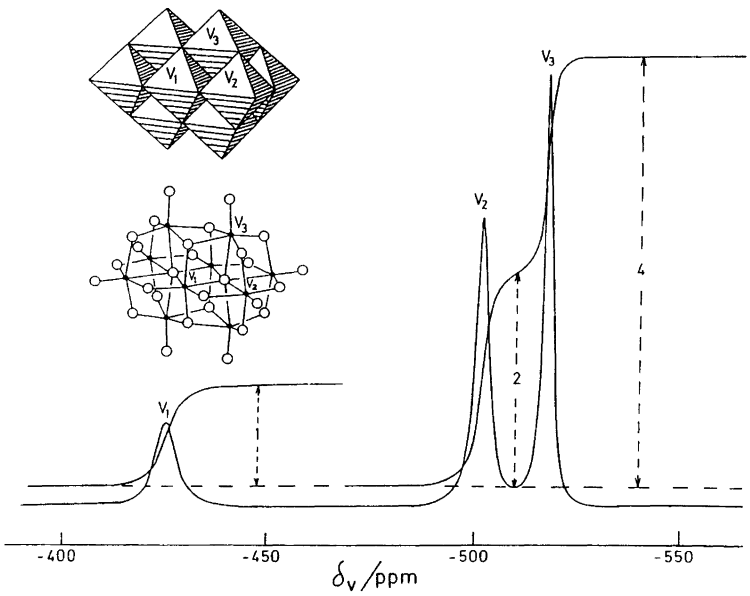


Fig. 1. ^{51}V NMR spectrum of the decavanadate unit $V_{10}O_{28}$. The integration of the peak areas gives the ratios $V_1:V_2:V_3=1:2:2$ (see text). The structure of the $V_{10}O_{28}$ unit is shown at the upper left hand part of the figure and the peaks arize from the three structurally different V-atoms in the unit.

ligand reactions according to eqn. (Va), (Vb) or (Vc), the peak position will be dependent on the free proton and/or ligand concentration.

7.2.2. The shift function. Fast exchange of protons and/or ligands according to the reactions (Va), (Vb) or (Vc) causes an averaging of individual shift values $\delta_{i,j}$ for all the (i,j)-species present. The relationship between the observed shift value, δ_{OBS}, and the species concentration, $cx_{i,j}$, is then given by

$$\delta_{OBS}=\sum\sum \delta_{i,j}cx_{i,j}/\sum\sum cx_{i,j} \qquad \text{(VI)}$$

where the summation is taken over all species present (including also the core species $i=0$, $j=0$). Since the core species is present in all the (i,j) species, the relation (VI) simplifies to

$$\delta_{OBS}=\frac{\sum\sum \delta_{i,j}k_{i,j}h^i l^j}{1+\sum\sum k_{i,j}h^i l^j} \qquad \text{(VIa)}$$

i.e. the shift values δ are independent of the total concentration B_{PQR} and are only a function of the proton and ligand concentrations h and l.

7.2.3. Shift profile applications. As seen from relation (VIa), the shift values and analysis of shift profiles directly give protonization and ligation functions in homonuclear series of complexes. Below we shortly describe two important applications:

(i) A shift curve can in principle be considered a formation curve and enables the determination number of proton steps (i,j) and corresponding step constants $(k_{i,j})$.

(ii) A *pqr*-search in an unknown multicomponent system is considerably simplified if protonization and/or ligation steps can be determined separately in some way. By using EMF data alone such information may often be difficult to obtain. A combined use of POT and NMR data offers unique possibilities, especially due to the fact that the NMR-peaks can independently give homonuclear proton or ligand equilibria and at the same time, from the peak integrals, subsets of equilibrium concentrations.

7.3. Additional possibilities by using the NMRI and NMRS routines.

7.3.1. Subsets of B_{tot}. Phase distribution data of the type DISTR, SURF, SOLID and GAS, see Table 4, can be treated exactly the same way as species distributions between peaks in an NMR spectrum (NMRI). The NMRI procedure can thus be used without any changes for these types of data.

7.3.2. Linear combinations of cx(i). It is possible to treat data of the type SPEFO, KALLE and CONDU using the NMRS procedure (Table 4). It is, however, necessary to divide observed absorbances, heats and conductivities with a $\Sigma\, cx(i)$ term, which can be the total component sum, B, or a subset of B. If these types of functions are plotted versus pH, one obtains 'Z-curves' of exactly the same type as the δ_{OBS}(pH)-plot given in Fig. 2.

8. Data selection and weightings

8.1. The data selection. In the LAKE system it is possible to select suitable data from the data base. These selection possibilities are a necessity for performing meaningful β-determinations in multimethod systems. The following types of selections can be used: (i) sets

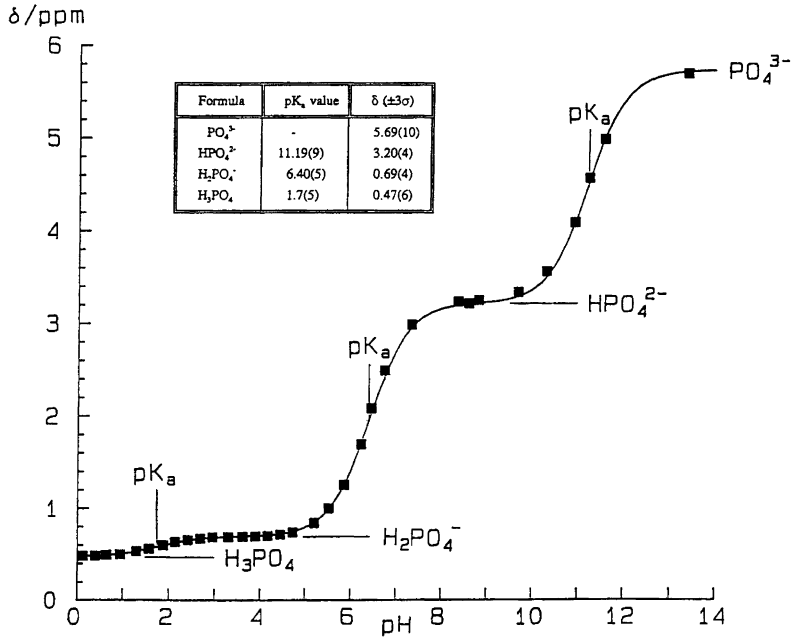


Fig. 2. ^{31}P chemical shifts as a function of pH for the H^+–PO_4^{3-} system. The symbols represent experimental NMR points and the curve has been calculated using the constants obtained in a LAKE calculation.

(intervals and specific sets), (ii) points, (iii) ranges of concentrations, (iv) ranges for the total concentration of different components, (v) ranges for quotients between total concentrations, and (vi) Z or $\bar{n}$ intervals.

8.2. The weighting. An important and difficult task in treating multimethod data is to give the different types of data their proper weights in the residual sum of squares (IV). In LAKE two types of weighting are used for the mass balance residuals. For each type of measured equilibrium quantity, individual weights are used. For NMRI data individual peak quality weights may be used (QUAW). The weighting possibilities are discussed below.

8.2.1. Residual weighting of the mass balances. The residual weighting for each mass balance equation is determined by setting weighting parameters w_j^i, $j=0$, ..., NCOM, $i=1$, ..., NCOM. By default $w_0^1=1$, $w_0^2=100$, $w_0^3=100$, ... and $w_j^i=0$, $i\geqslant 1, j\geqslant 1$. There are two principal cases, the pure Total concentration weighting and a generalized Z and $\bar{n}$ type of weighting.

(i) Total concentration weighting. If only w_0^i, $i=1$, ..., NCOM, have nonzero values LAKE defines the mass balance residuals as $\Sigma_{i=1}^{\mathrm{NCOM}}(w_0^i\, r_i^{\mathrm{MBE}})^2$. In LETAGROP all errors are placed on the first equation, H. The components B, C, ... are assumed correct. The default weighting gives similar results to LETAGROP.

(ii) Generalized Z and n̄ type of weighting. If for some $i\geqslant 1, j\geqslant 1$, it holds that $w_j^i>0$, then LAKE uses a more general weighting formula and defines the mass balance residual sum of squares as

$$\sum_{i=1}^{\mathrm{NCOM}}\left(\frac{w_0^i}{w_1^i[\mathrm{H}]+w_2^i[\mathrm{B}]+w_3^i[\mathrm{C}]+\ \ldots}\, r_i^{\mathrm{MBE}}\right)^2$$

Any zero denominator occuring in this expression for some component i is replaced by the value one, i.e. the weighting for this component then is of the form used in (i).

In equilibrium analysis the quantity Z, i.e. the average number of protons bound per B, C, ..., or some suitable sum of these is commonly used. $\bar{n}$, the average number of ligands bound per metal ion is also a very useful function. Generally these relative functions give more precise chemical information than the total concentrations. In binary systems particularly, $Z(\mathrm{pH})$ curves at constant B are very instructive by giving the first hint of the speciation pattern. Such a set of data is often the starting point in equilibrium analysis. In a potentiometric titration where $-\log h$ is measured, Z can directly be obtained from the experimental data by using the relationship $Z=(H-h+k_w h^{-1})/B$, where $h=[\mathrm{H}^+]$ and $k_w=$ the ionic product of water. Giving values to the weights w_1^i, w_2^i, ... means dividing the ith mass balance with a linear combination of the total concentrations and gives generalized Z and $\bar{n}$ functions to be used in the optimization procedure.

8.2.2. Residual weighting of the measured equilibrium quantities. Besides the described weighting of the total concentrations it is also possible to define residual weights for all additional types of data using the program DEFW.

For EMF (electrode measurements) and POT

($-\log\{\ \}$-measurements) the error squares sums are defined as

$$U_{\mathrm{EMF}} \sum_{i=1}^{\mathrm{NELEC}} (w_i^{\mathrm{EMF}} r_i^{\mathrm{EMF}})^2 \text{ and } U_{\mathrm{POT}} = \sum_{i=1}^{\mathrm{NLOG}} (w_i^{\mathrm{POT}} r_i^{\mathrm{POT}})^2$$

NELEC and NLOG are the number of electrodes and $-\log\{\ \}$ measurements. Correspondingly for NMRI and NMRS

$$U_{\mathrm{NMRI}} = \sum_{i=1}^{\mathrm{NNUCL}} \sum_{j=1}^{\mathrm{NTOP}} (w_{i,j}^{\mathrm{NMRI}} (\mathrm{QUAW})_{i,j} r_{i,j}^{\mathrm{NMRI}})^2 \text{ and}$$

$$U_{\mathrm{NMRS}} = \sum_{i=1}^{\mathrm{NNUCL}} \sum_{j=1}^{\mathrm{NTOP}} (w_{i,j}^{\mathrm{NMRS}} r_{i,j}^{\mathrm{NMRS}})^2$$

NNUCL and NTOP are the number of nuclei and peaks, respectively. For NMRI, the peak quality weighting QUAW is used. The user defines n_q quality groups/classes and assigns a weight to each class. Each NMRI measurement is assigned in the input file to belong to one of these n_q quality classes.

The EMF residual is mainly used when estimating special parameters in the electrode equation, i.e. E_0, k_{EL}, formation constants of dirt acids and liquid junction potentials. All residual types except the MBE-residual, r^{MBE}, may freely be included or excluded in a given calculation whereas r^{MBE} is always included.

9. Case studies

Two case studies will be shown. Case 1 is a simple case, the binary phosphate system, where only a series of mononuclear proton complexes are formed, and illustrates the use of NMR chemical shift data (NMRS). Case 2 exemplifies a much more complicated case, the ternary molybdophosphate system, where combined potentiometric data (EMF, POT) and NMR data (NMRI, NMRS) have been used to determine the speciation.

9.1. Case 1; the H^+–PO_4^{3-} system – shift profile curves. The simple binary system H^+–PO_4^{3-} is made up of a series of mononuclear proton complexes $H_p(PO_4^{3-})$ with $p = 0$, 1, 2 and 3. These compositions are well established from potentiometric data. On the NMR time scale, proton reactions are too rapid for detecting individual species and hence only one peak is present in the ^{31}P NMR spectra. The peak has, however, a marked pH dependence as shown in Fig. 2. The curve shows three steps of which two are distinct. The plateaus represent the range in which either of the species is predominant and the inflection points represent the pk_a-values. The data are taken from a study of the vanadophosphate system in 0.6 M Na(Cl) medium at 25 °C.[34]

From the calculation on (NMRS+POT)-data (43 experimental points) the protonation and the shift constants given in the table inserted in Fig. 2 were obtained. As expected, the pk_a-values for $H_2PO_4^-$ and HPO_4^{2-} are well determined (the two distinct steps in the shift curve) in contrast to that for H_3PO_4 (the small step). To obtain more accurate values potentiometric titration data (185 experimental points) were included. $H_2PO_4^-$ was chosen as component and calculations on the combined NMRS+POT and EMF data resulted in the values shown in Table 6. By simultaneous treatment of combined data, all formation constants could thus be determined with high precision.

9.2. Case 2; the H^+–MoO_4^{2-}–HPO_4^{2-} system – combined EMF, POT and NMR data. A potentiometric and ^{31}P NMR equilibrium study of the molybdophosphate system had earlier been performed in 3.0 M $Na(ClO_4)$ medium.[35,36] A complementary study has recently been performed in 0.6 M Na(Cl) medium. Most spectra with overlapping peaks have now been recorded on a 500 MHz spectrometer, and LAKE has been used for establishing the speciation. The full study will be presented elsewhere.[32] In this chapter some parts of the study will be related, primarily to show how LAKE has been used for establishing the speciation. The following data types have been collected: EMF (potentiometric titration data), POT (potentiometric point data), NMRI+POT (peak integral data) and NMRS+POT (chemical shift data).

9.2.1. Core units in the system – analysis using shift profiles. Since the shift data (NMRS) were not found to be very decisive, these data were not included in the LAKE calculations. However, the shift profile curves give valuable information about protonation of the different core units. In Fig. 3 the ^{31}P shift values are plotted as a function of pH. Besides the peak arizing from the binary H^+–HPO_4^{2-} species [(0,1) in (q,r)-notation)], resonances from eight other 'core' units are present and most of these take part in proton equilibria. The total number of species is thus considerable and hence the system is very complicated from an equilibrium analysis point of view.

9.2.2. The (5,2) core unit – analysis using the (p,q,r)-grid. At Mo/P ratios less than 2.5 only one ternary peak is present except for in very acid solutions. However, the peak position is pH dependent indicating protonation of the core unit. In order to find the composition of the first member in the series a (p,q,r)-analysis was performed. The data selection facility (section 8.1) was used to restrict the pH-range ($5.5 < \mathrm{pH} < 7.0$) and the Mo/P ratio ($\mathrm{Mo/P} \leqslant 2.5$). Only nine EMF points and seven NMR points fulfilled this criteria. On these limited data, a (p,q,r)-search was performed with the grid facility. A

Table 6. Formation and chemical-shift constants for the binary H^+–$H_2PO_4^-$ system obtained from a calculation on combined NMRS and EMF data.[34]

(p,q,r)	Formula	$\log\beta(\pm3\sigma)$	pK_a-value	$\delta\ (\pm3\sigma)$
−2,0,1	PO_4^{3-}	−17.649(6)	–	5.73(5)
−1,0,1	HPO_4^{2-}	−6.418(4)	11.231	3.22(2)
0,0,1	$H_2PO_4^-$	0.	6.418	0.70(2)
1,0,1	H_3PO_4	1.772(3)	1.772	0.48(3)

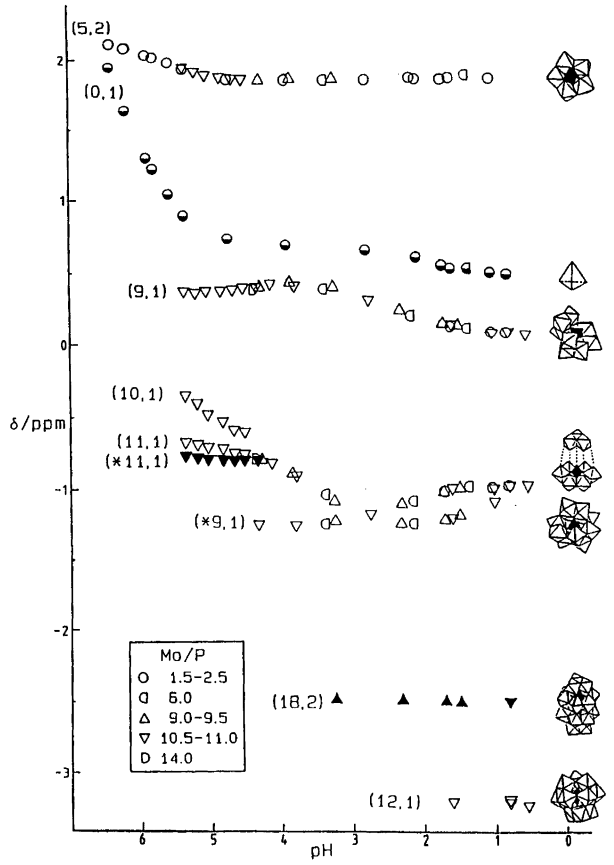


Fig. 3. Experimental ^{31}P chemical-shift values plotted as a function of pH for all different $(q,r)=Mo_qP_r$ 'cores' of molybdophosphate species.[32]

(p,q)-grid with half unit steps was calculated for $r=1$, 2 and 3. The lowest U-value (error squares sum) was obtained for $r=2$, and this grid is presented in Fig. 4a. As seen in the figure $p=8.0$ and $q=5.0$, and thus the composition $(H^+)_8(MoO_4^{2-})_5(HPO_4^{2-})_2$ gave the 'best' explanation. As seen from the U-values the pit does not have the shape of a centric hole, but rather that of an ellipse as indicated in the figure. To further examine the bottom of the pit, U-values were calculated with 0.1 steps in p and q and for a smaller region (dashed frame, Fig. 4a). Then isolines for constant U-values beginning at 6×10^{-4} and in steps of 2×10^{-4} were drawn as shown in Fig. 4b. The isolines are indeed elliptical. The U-surface has the shape of a groove characterized by a p/q slope ca. 1.6 (=8/5) and with the lowest value, $U_{min}=5.82\times10^{-4}$ for $p=8.0$ and $q=5.0$.

A $\beta_{8,5,2}$ calculation was then performed in order to examine the optimization and residual reports and to check if the weightings used seem reasonable.

The Global Optimization Report, Table 7, exemplifies the result of a log β determination using combined EMF, POT and NMRI data. The optimization implies a minimization of a sum, U_{SUM}, of weighted squared residuals $\sum(wr)^2$. In the present case there are three residual sums for the mass balances MBE(H), MBE(B), MBE(C), and two for the peak integrals NMR(1) and NMR(2) corresponding to the ternary and binary peaks, respectively. The mass balance equations have been given the default weights 1, 100, 100, and the NMR peaks the weights 10, 10.

The Residual Statistics Report, in Table 7, based on the log β_{pqr} optimization displays squared sums of weighted residuals, $\sum[w(y_{calc}-y_{obs})]^2$ for the different types of residuals used. In addition, contributions (in %) to U_{SUM} for the different residual types are calculated.

Note that the total number of residuals is $\sum n_{Type}=62$ $(3\times16+2\times7)$ as seen in the Residual Statistics Report. The separable least-squares algorithm does a projection, which reduces the number of residuals in the global optimization with the number of local parameters estimated; in this case the free concentrations b and c. Reducing 62 with 32 (2×16) gives the residual count 30 in the Global Optimization Report.

Since the pH-range is very small, the EMF data should not be very decisive. Therefore, the distribution of 14% from MBE and 86% from NMR was considered reasonable, as well as the $3\sigma[\log\beta(8,5,2)]$-value of 0.09.

The pH-range was then extended to more acid solutions (pH<2). Protonated pentamolybdodiphosphate species were included in the model and the series (8,5,2), (9,5,2) and (10,5,2) completely explained all EMF and NMRI data. Figure 5 shows that the fit of the experimental NMR data to the found model is very good. Since both of the two peaks are composed of a series of homonuclear species, the solid lines in the figure have been calculated as the sum of all the (0,1) and the (5,2) species, respectively. The dashed curves show the distribution of the individual $(p,5,2)$ species. In our opinion it is extremely important to compare the proposed model with experimental data as shown in Fig. 5. The distribution diagram has been calculated using the SOLGASWATER program.[37] However, a plotting procedure for LAKE is currently being developed, which will aid in the search for finding the 'best' model.

9.2.3. The core units (9,1), (10,1) and (11,1) – example with isomers. The Mo–P system becomes much more complicated when the Mo/P ratio is increased (>2.5). At ca. pH 5 four new peaks denoted (9,1), (10,1), (11,1) and (*11,1) appear (Fig. 3). The asterisk is used to denote isomeric species. At somewhat lower pH-values a fifth peak appears, (*9,1). Since these five peaks arise from species with similar Mo/P ratios. EMF data alone are not decisive enough. The H–Mo–P system is indeed suitable for a LAKE treatment since high precision combined NMR–EMF data are necessary to solve the speciation. Four of the peaks have overlapping regions, and the NMRi program[38] had to be used for the quantitative evaluation of most spectra. After an extensive data treatment using LAKE, where again the selection facility was widely used, the model given in Table 3 for the $H_p(MoO_4)_q(HPO_4)_r$ species was established. Their pH-range of existence is shown in Fig. 3.

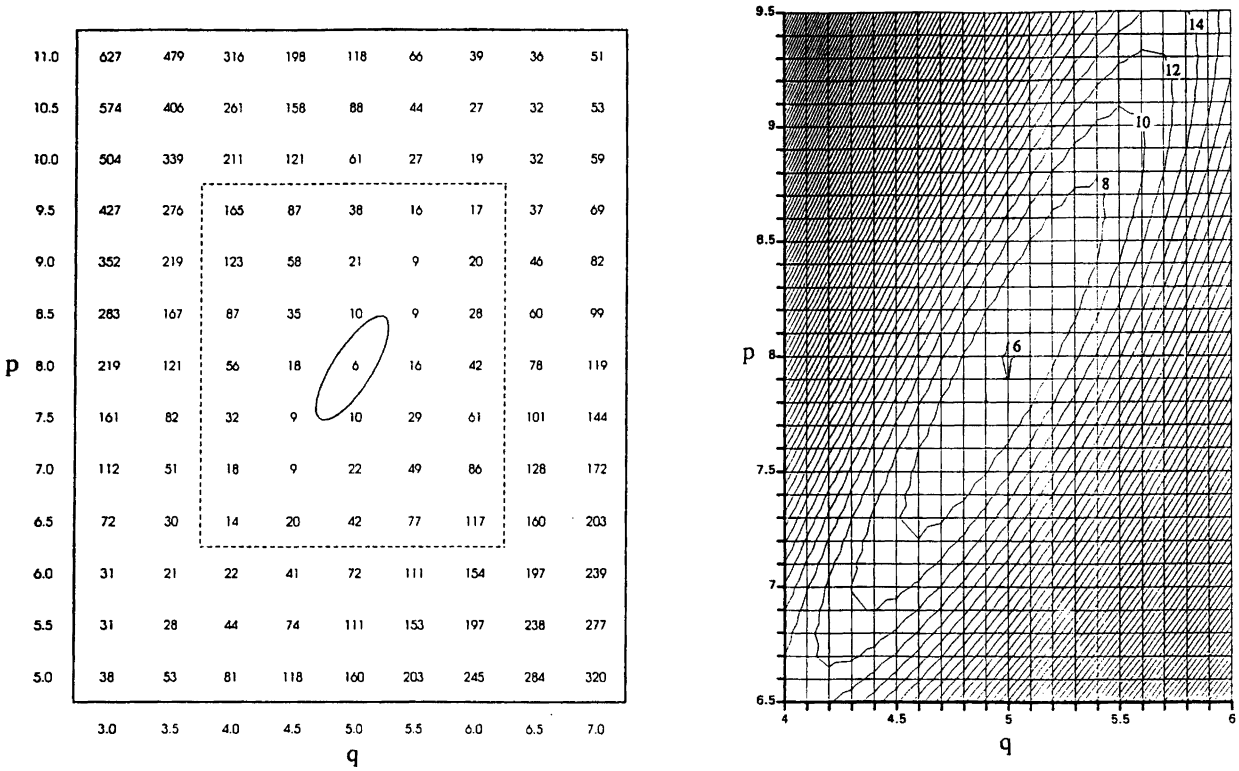


Fig. 4. (a) A (p,q)-grid for $r=2$ in the pH-range $5.5<\text{pH}<7.0$ and at Mo/P $\leqslant 2.5$. (b) Isolines for constant U-values from the dashed framed area in (a).

Table 7. Optimization and statistics reports from a $\beta_{8,5,2}$ calculation in the H^+–MoO_4^{2-}–HPO_4^{2-} system.

Global Optimization Report

Significant variables	1
Optimized variables	1
Residuals	30
U_{MIN}	5.8212×10^{-4}
$[U_{MIN}/(n-1)]^{1/2}$	44.8030×10^{-4}

cx(i)	*pqr*	Starting value (log)	Optimal constant (log)	3σ (log β_{pqr})
13	8,5,2	57.3800	57.5248	0.0865

Residual Statistics Report

Type	n_{Type}	w_{Type}	U_{Type} ($\times10^{-4}$)	% of U_{SUM}
MBE(H)	16	1	0.7864	13.51
MBE(B)	16	100	0.0056	0.10
MBE(C)	16	100	0.0138	0.24
NMR(1)	7	10	2.5258	43.39
NMR(2)	7	10	2.4785	42.77
U_{MBE}			0.8054	13.84
U_{NMR}			5.0154	86.16
U_{SUM}			5.8212	

9.2.4. The core units (12,1) and (18,2). In more acidic high ratio Mo/P solutions an additional species is formed having a (12,1) composition and likely having the so-called α-Keggin structure shown at the bottom right of Fig. 3. Minor amounts of an (18,2) dimer is also slowly formed and its formation constant has been determined from aged NMR solutions.

9.2.5. Concluding remarks. Most of the structures shown in Fig. 3 have been determined by single crystal

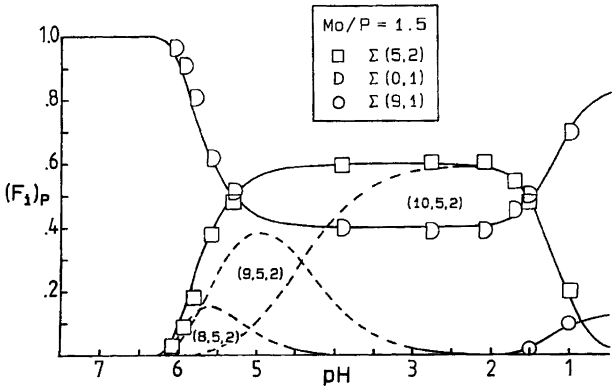


Fig. 5. Diagram showing the distribution of phosphorus as a function of pH at Mo = 30 mM and P = 20 mM. The symbols represent experimental NMR points. The full and dashed curves have been calculated using the model given in Table 3.

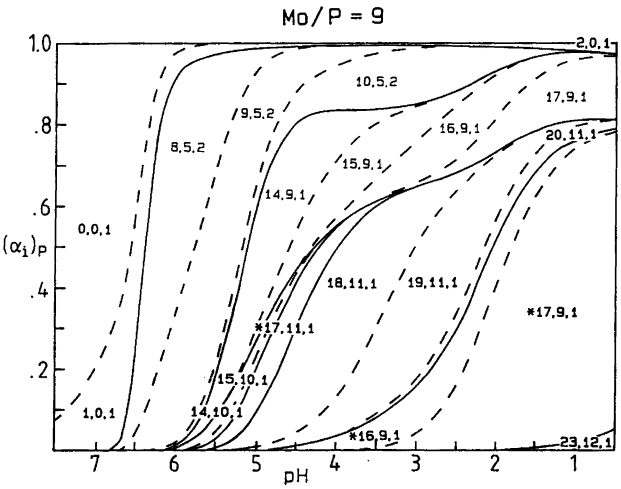


Fig. 6. Diagram showing the cumulative distribution of phosphorus containing species as a function of pH at Mo = 180 mM, P = 20 mM. Each (*p,q,r*) species is represented by an area and its fraction at a given pH is the vertical height of its area at that pH. The curves have been calculated using the model given in Table 3. Full curves separate species with different, and dashed curves species with the same 'core' unit.

X-ray methods and some of the structures verified by large-angle X-ray studies on concentrated solutions. These studies are partly discussed in Ref. 36. Some of the structures are only proposed, such as those for the (9,1) and (11,1) series. Besides the (18,2) species, which is present only in aged solutions, 13 different species having a Mo/P ratio between 9 and 12 are co-existing. Figure 6 represents a type of diagram which clearly shows the distribution even in such complex systems as the present one. The diagram has been calculated with the SOLGASWATER program[37] and shows the distribution of P-containing species at Mo/P = 9. This diagram does not show the entire complexity of the molybdophosphate system, since the distribution to binary H^+–MoO_4^{2-} species cannot be shown in a P-diagram. The final equilibrium model presented in Table 3 could never have been established without quantitative NMR-data and the LAKE program.

Acknowledgements. We are grateful to The Swedish Natural Science Research Council (NFR) for financial support. We also thank Dr. Justin Powlowski for linguistic revision.

References

1. Ingri, N. and Sillén, L. G. *Ark. Kem. 23* (1964) 97.
2. Sabatini, A., Vacca, A. and Gans, P. *Talanta 21* (1974) 53.
3. Gans, P., Sabatini, A. and Vacca, *A. J. Chem. Soc., Dalton Trans.* (1985) 1195.
4. Sabatini, A. and Vacca, A. *Coord. Chem. Rev. 120* (1992) 389.
5. Meloun, M., Havel, J. and Högfeldt, E. *Computation of Solution Equilibria: A Guide to Methods in Potentiometry, Extraction and Spectrophotometry.* Ellis Horwood Ltd., New York 1988.
6. Leggett, D. J. (ed.) *Computational Methods for the Determination of Formation Constants*, Plenum Press, New York 1985.
7. Holmström, K. Doctoral thesis, University of Umeå, Umeå, Sweden 1988.
8. Gans, P., *Data Fitting in the Chemical Sciences by the Method of Least Squares*, John Wiley & Sons Ltd, Chichester 1992.
9. Dyrssen, D., Ingri, N. and Sillén, L.G. *Acta Chem. Scand. 15* (1961) 694.
10. Sillén, L. G. *Acta Chem. Scand. 16* (1962) 159.
11. Sillén, L. G. and Warnqvist, B. *Ark. Kem. 31* (1969) 341.
12. Ingri, N. and Sillén, L. G. *Acta Chem. Scand. 16* (1962) 173.
13. Brauner, P., Sillén, L. G. and Whiteker, R. *Ark. Kem. 31* (1969) 365.
14. Sillén, L. G. and Warnqvist, B. *Ark. Kem. 31* (1969) 377.
15. Arnek, R. *Ark. Kem. 32* (1970) 81.
16. Liem, D. H. *Acta Chem. Scand. 25* (1971) 1521.
17. Warnqvist, B. *Chem. Scr. 1* (1971) 49.
18. Pettersson, L., Hedman, B., Andersson, I. and Ingri, N. *Chem. Scr. 22* (1983) 254.
19. Sjöberg, S., Öhman, L.-O. and Ingri, N. *Acta Chem. Scand., Ser. A 39* (1985) 93.
20. Schindler, P. and Gamsjäger, H. *Kolloid Z. Z. Polym. 250* (1972) 759.
21. Andersson, L., Holmström, K. and Ruhe, A. *Algorithms for Approximation*, Vol 10, pp. 557–572, Oxford University Press, London 1987.
22. Andersson, L. and Holmström, K. *Report UMINF-142.88.* Institute of Information Processing, University of Umeå, S-901 87 Umeå, Sweden.
23. Ruhe, A. and Wedin, P.-Å. *SIAM Rev. 22* (1980) 318.
24. Dongarra, J. J., Moier, C. B., Bunch, J. R. and Stewart, G. W. *LINPACK User's Guide*, SIAM, Philadelphia 1979.
25. Lindström, P. Doctoral Thesis, University of Umeå, Umeå, Sweden 1983.
26. Fletcher, R. *Practical Methods of Optimization. Vol. 1: Unconstrained Optimization*, John Wiley, New York 1980.
27. Ljung, L. *System Identification*, Prentice-Hall, Englewood Cliffs, NJ 1987.
28. Pettersson, L. In Pope, M. T. and Müller, A., Eds., *Polyoxometalates: From Platonic Solids to Anti-retroviral Activity*, Kluwer Academic Publishers, Dordrecht, The Netherlands 1994, pp. 27–40.
29. Yagasaki, A. Doctoral Thesis, The University of Tokyo, Tokyo 1986.
30. Yagasaki, A., Andersson, I. and Pettersson, L. *Inorg. Chem. 26* (1987) 3926.

31. Howarth, O. W., Pettersson, L. and Andersson, I. *J. Chem. Soc., Dalton Trans.* (1991) 1799.
32. Pettersson, L. and Andersson, I. *To be published.*
33. Blixt, J., Györi, B. and Glaser, J. *J. Am. Chem. Soc. 111* (1989) 7784.
34. Selling, A., Andersson, I., Pettersson, L., Schramm, C. M., Downey, S. L. and Grate, J. H. *Inorg. Chem. 33* (1994) 3141.
35. Pettersson, L., Andersson, I. and Öhman, L.-O. *Acta Chem. Scand., Ser. A 39* (1985) 53.
36. Pettersson, L., Andersson, I. and Öhman, L.-O. *Inorg. Chem. 25* (1986) 4726.
37. Eriksson, G. *Anal. Chim. Acta 112* (1979) 375.
38. Dumoulin, C. L. and Levy, G. C. *J. Mol Struct. 113* (1984) 299.

Received 5 July, 1995.