On Equilibria in Systems with Polynuclear Complex Formation. I. Methods for Deducing the Composition of the Complexes from Experimental Data. "Core + Links" Complexes

LARS GUNNAR SILLÉN

Department of Inorganic Chemistry, Royal Institute of Technology, Stockholm 70, Sweden

The formation of a series of consecutive *mononuclear* complexes has been treated by a number of authors. Equations and diagrams have been given to show how the equilibria are shifted with varying composition of the solution and how the equilibrium constants involved can be calculated from experimental data of varying type and accuracy (*e.g.* N. Bjerrum [1], J. Bjerrum [2], Leden [3]).

The formation of *polynuclear* complexes has not met with the same interest. In general they have been considered as an annoying disturbance, impeding the measurement of the interesting mononuclear equilibria, and not being studied for their own sake.

Sometimes, as for instance in the chromium(III) hydrolysis studied by N. Bjerrum and Faurholt [4,5], the polynuclear complexes are formed by very slow reactions. In other systems, such as some of the copper-complex systems studied by Fronaeus [6], the polynuclear complexes seem to be formed by rapid reversible equilibria. Fronaeus eliminated them by extrapolation of his data to low concentrations, where the mononuclear complexes predominate. A similar procedure was attempted in Ahrland's paper [7] on the hydrolysis of the uranyl ion, although the main reaction products in this case are obviously polynuclear complexes.

In systems where polynuclear complexes occur, the equilibrium data have sometimes been explained in terms of reactions by which only one or two polynuclear species are formed. Such an explanation may, in a certain concentration interval, be approximately true for the equilibria between CrO_4^{2-}, $HCrO_4^-$, and $Cr_2O_7^{2-}$ in dilute acid chromate solutions (see *e.g.* Neuss and Rieman [8]), although there is certainly evidence for the formation of other polynuclear complexes, *e.g.* $Cr_3O_{10}^{2-}$. In many other cases one may doubt whether it is experimentally proved or even at all likely that Nature would select, out of all possible polynuclear complexes, exactly the one or two proposed by a certain chemist.

In later years, evidence has accumulated that in many systems an apparently unlimited series of polynuclear complexes are formed. It is believed that the first attempt to formulate such a mechanism quantitatively was the work of Granér and Sillén [9] on the hydrolysis of the ion Bi^{3+} which could be explained in terms of the formation of polynuclear complexes $Bi_{n+1}O_n{}^{(n+3)+}$. Later on, Brosset [10] has suggested that the aluminium ion forms complexes $Al_n(OH)_{3(n-1)}^{3+}$ in the acid pH range and $Al_n(OH)_{3n+1}^-$ in the alkaline range. Recently, Leden [11] has found evidence for the formation of complexes $Ag_nI_{2n+2}^{(n+2)-}$, and the corresponding complexes of thiocyanate and silver ions. As will be shown in subsequent papers, a similar mechanism can explain the equilibrium data in several other systems as well.

If polynuclear complexes are formed, the number of possible complexes, and thus of equilibrium constants, is infinitely larger than when there are only mononuclear complexes. If one finds one mechanism that can explain the experimental data, one may thus be in doubt whether the same data could not be explained equally well by a quite different mechanism.

In the present paper and a few ones following an attempt will be made to discuss the mathematical laws for the formation of polynuclear complexes, both under rather general assumptions and for a few special mechanisms. The chief aim has been to find out how much information on the mechanism one can really find from the experimental data, which by necessity are of limited accuracy and cover a limited range of concentrations.

The discussion is not limited to complexes of a metal ion and some ligand. The two complex-forming species A and B in the following treatment may be any charged or uncharged molecular species: Ag^+ and I^- or I^- and Ag^+, a metal ion and OH^- or any other ligand, a fatty acid HA and the corresponding ion A^-, H^+ and A^- or SiO_2 and OH^- etc.

In this paper we have made two fundamental assumptions: *i.e.* that we are dealing with real equilibria and that all complexes are dissolved in the solution. If equilibrium has not been attained, all arguments based on the law of mass action are of course invalid unless the slow reactions are so slow that they can be neglected.

The precipitation of a solid phase of constant composition will be found as a special case of several mechanisms discussed in the following (see part II) [12]. However, there is also the possibility that A and B form a separate phase of composition variable within a certain range. This behavior may not be expected to be usual for the hydrolysis products of metal ions or for the polynuclear silver halogenide complexes but has been proposed for the micelles in soap solutions. We hope to be able to return to this point later on in connection with work now in progress in this institution.

Equations for equilibria with a solid solution phase are not very difficult to set down. However, it is felt that one should first have real systems to test the equations in order to see which of them are really practicable with the present accuracy of measurement.

List of symbols

A total concentration of A (2)

a concentration of free A (1)

B total concentration of B (2)

b concentration of free B (1)

C $= B =$ "total concentration of C" (32)

c free concentration of complex $C = A_{r+t}B$ (30)

c_n concentration of n:th complex in (21) or (31)

f $= Sb^{-1}$, function of u (22)

k_n equilibrium constant, defined by (21) or (31)

N singular value for n, text around (27 to 29)

n variable integer; number of links in a "core + links" complex

P singular value for p, (27 to 29)

p number of A in a complex (2)

Q singular value for q (27 to 29)

q number of B in a complex (2)

R constant partial derivative in (16), obtained from spacing between $Z(\log a)_B$ curves

r constant number of A in core (if free from B); (8 a) and Table 1

S complexity sum (4)

s constant number of B in core (if free from A); (8 a) and Table 1

T constant partial derivative in (40), obtained from spacing between $\log \dfrac{B}{b}$ $(\log a)_B$ curves

t constant number of A per B in link; (8 a) and Table 1

u $= a^t b$ (10) $(= a^{-r} c$ in (30))

x $= t \log a + \log B$ (26)

y $= \dfrac{A-a}{Bt} = \dfrac{Z}{t}$ (25)

Z $= \dfrac{A-a}{B}$ (7), average number of A bound per B

η $= \log \dfrac{B}{b}$ (44)

$\varkappa_n$ equilibrium constant, defined by (9)

$\varkappa_{pq}$ equilibrium constant defined by (3)

φ $= Sa^{-r}$, function of u (11)

General equations

We shall denote the two reactants by A and B. They may be charged but for simplicity no charges will be indicated. Let the total concentrations of the reactants, including the amounts bound in complexes, be A and B mole 1^{-1}, and the free concentrations.

$$[A] = a \text{ and } [B] = b \tag{1}$$

When A and B react, one or several complexes A_pB_q are formed in the solution; here p and q may be any integer. For the total concentrations A and B we have:

$$A = a + \Sigma p[A_pB_q]; \quad B = b + \Sigma q[A_pB_q] \tag{2}$$

where the summations are made over all possible sets (p, q) excluding $(1, 0)$ and $(0,1)$ which correspond to free A and B.

We shall assume that we can regard the activity factors as constant; if A or B or both are charged, the reactions must be studied in a constant ionic medium. Consequently we can use concentrations instead of activities in the law of mass action and measure ionic concentrations directly by emf methods.

For the formation of any complex A_pB_q we then have

$$p\text{A} + q\text{B} \rightleftharpoons A_pB_q; \quad [A_pB_q] = \varkappa_{pq}a^pb^q, \tag{3}$$
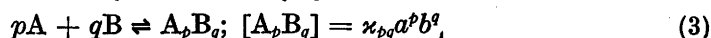
where $\varkappa_{pq}$ is the equilibrium constant of the reaction.

It is convenient to introduce a quantity S which might be called the *complexity sum;*

$$S = \Sigma[A_pB_q] = \Sigma\varkappa_{pq}a^pb^q \tag{4}$$

From equations (1—4) we find the amount of A bound in complexes, $A—a$ mole/1:

$$A—a = \Sigma p[A_pB_q] = \Sigma p\varkappa_{pq}a^pb^q = a\left(\frac{\partial S}{\partial a}\right)_b = \left(\frac{\partial S}{\partial \ln a}\right)_b \tag{5}$$

Similarly the amount of B bound in complexes, $B—b$ mole B/1, is

$$B—b = \Sigma q\varkappa_{pq}a^pb^q = b\left(\frac{\partial S}{\partial b}\right)_a = \left(\frac{\partial S}{\partial \ln b}\right)_a \tag{6}$$

A quantity which we shall often have use for is the average number of A bound per B; we shall denote it by Z

$$Z = \frac{A—a}{B} \tag{7}$$

It follows from (7), (5), and (6)

$$Z = \frac{a\left(\frac{\partial S}{\partial a}\right)_b}{b\left[1 + \left(\frac{\partial S}{\partial b}\right)_a\right]} = \frac{\left(\frac{\partial(S + b)}{\partial \ln a}\right)_b}{\left(\frac{\partial(S + b)}{\partial \ln b}\right)_a} \tag{7 a}$$

Types of experimental data

In studies on complexity equilibria, it should always be possible to measure the total concentrations A and B by chemical analysis or to calculate them from the way the solution has been prepared. It is often convenient to let one of them, say B, be constant in each group of experiments. Then B has only a limited number of values: B_1, B_2, B_3, etc., whereas A is varied freely.

It is often possible to measure one of the free concentrations a and b, sometimes both. Since the following equations will be applicable even if A and B are interchanged, we may restrict our treatment to two cases:

Data (a, A, B). (Equations 14—38). Only a is measured but not b. The results may conveniently be plotted by giving $Z = \dfrac{A-a}{B}$ as a function of $\log a$, with one curve for each B value: B_1, B_2 etc. An example is the hydrolysis of the thorium ion (Hietanen [13]) where [H^+], which is proportional to [OH^-]$^{-1}$, was measured by emf methods but not [Th^{4+}].

Data (a, b, A, B). (Equations 39—49) Both a and b are measured. If, say, b is more accurately known than a, then $\dfrac{b}{B}$ and $\dfrac{B-b}{A}$ are the most accurately known functions but $\dfrac{A-a}{B}$ and $\dfrac{A-a}{B-b}$ may also be calculated and plotted, *e.g.* as functions of $\log a$. An example is the hydrolysis of Bi^{3+} where both [Bi^{3+}] and [H^+] could be measured [9].

Other types of data: Suppose that one has measured the *solubility* of a certain solid A_PB_Q (which may in addition to A and B contain solvent molecules and ions from the ionic medium) *. Then one knows that in all solutions the product a^Pb^Q is a constant, at first unknown. In general, solubility data alone are insufficient for a mathematical analysis of the equilibria in a system with polynuclear complexes, since such data are limited to a single curve in the (a, b) field. With emf methods, on the other hand, one can cover a two-dimensional area.

Two-dimensional areas in the *(a, b)* field are covered also by *extraction* or *ion exchange* measurements. These methods have proved valuable for studying the formation of mononuclear complexes, in which only one uncharged complex is extracted (Rydberg [14]) or only one or a few charged complexes are taken up by the ion exchanger (Fronaeus [15]). If polynuclear complexes are formed, one may expect a number of different complexes to be extracted, or sorbed in the ion exchanger. Such data may be attacked by the same general methods as given below, after introducing a few more reasonable assumptions. They shall not be discussed here, however, nor shall we discuss optical data which do not seem suitable for investigating polynuclear complexes (*cf.* Fronaeus [6], p. 89 ff) unless it is possible to devise an optical method which gives directly the number of A—B bonds.

Representation of *(p, q)* set

The set of all pairs of non-negative numbers (p, q) can be represented as a two-dimensional array of points in one fourth of an infinite plane **. (Fig. 1) We shall denote by black dots ● those points corresponding to complexes

---

* *E.g.*, if A = I^-, B = Ag^+, and the ionic medium is $NaClO_4$ in water, then $NaAgI_2(H_2O)_{3\frac{1}{2}}(s)$ would count as $A_2B(s)$.

** If one would prefer to define the reactants so that p or q may become negative (say, to consider the formation of OH^- complexes as a subtraction of H^+), the (p, q) domain would be different, but nothing would be changed in the following argument and equations.
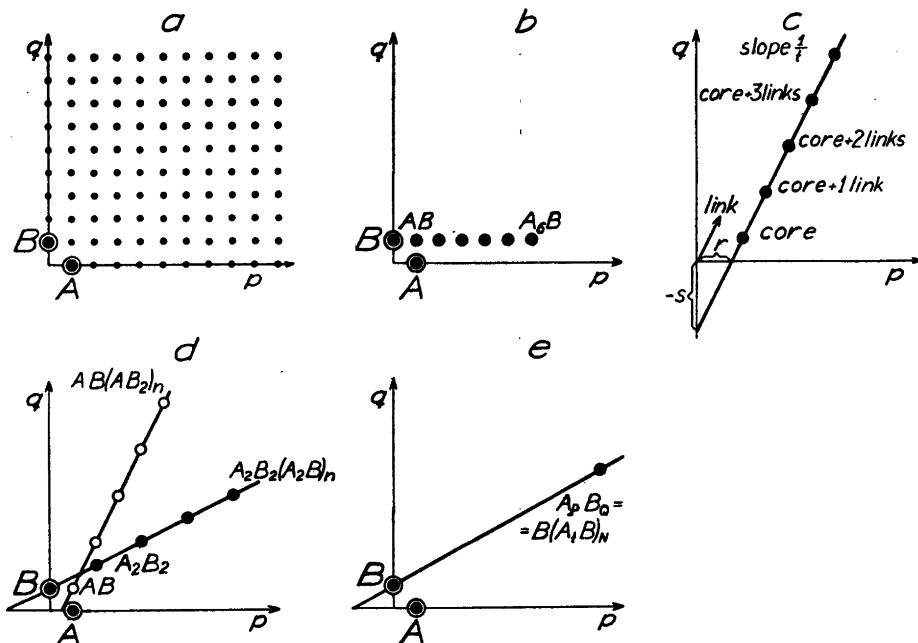
Fig. 1. Geometrical representation of the (p, q) set. Points = pairs of integers (p, q), dots¸ ● = pairs (p, q) corresponding to an existing complex A_pB_q. The pure components A and B are dots with circles ⊙ in a, b, d, and e. a) No complexes formed. b) Only mono-nuclear complexes A_nB (n = 1,2 ... 6). c) Compositions of complexes correspond to "core" + n "links"; the corresponding points, are on a straight line with intercepts r and s, and slope 1/t. d) Two "core + links" mechanisms: complexes $A_2B_2(A_2B)_n$ (r = -2, s = 1, t = 2; dots ●) and complexes $AB(AB_2)_n$ (r = ½, s = -1, t = ½; circles ○). e) Only one complex formed, $A_PB_Q = B(A_lB)_N$.

that actually occur in the solution. Fig 1 a represents a system where no com-plexes are formed: there are dots only for A (1,0) and for B (0,1). In Fig. 1 b there are in addition mononuclear complexes A_nB up to $n = 6$.

For any pair of reactants, it would seem unlikely that the polynuclear complexes, if they occur at all, would be spread equally over the whole area. For instance, if A and B are ions, complexes corresponding to points close to the coordinate axes are not likely to occur at a distance from the origin, since they would consist of large clusters of ions of the same sign, containing only a few opposite ions. Let us imagine that there is over each point (p, q) in the diagram, a rod, perpendicular to the paper, the height of which gives the concentration of the corresponding complex A_pB_q in a certain solution. Then it seems likely that the ends of the rods would lie on a ridge, making an angle with both coordinate axes.

For any type of data it is not so hard to devise some straight-forward mathematical method — involving, for instance, repeated division by a or b and extrapolation to 0, or repeated differentiation — that would finally give

us a complete set of equilibrium constants $\varkappa_{pq}$, one for each complex A_pB_q. Unfortunately such methods of calculation would require an accuracy for the data which is hardly ever met with in practice. So one must in practice test various simplifying assumptions.

THE "CORE + LINKS" HYPOTHESIS

Restrictions on p and q

We shall see in the following that a rather frequent type of experimental data leads to the conclusion that all complexes that exist in appreciable amounts have a composition corresponding to a certain "core" + a varying number of "links".

We shall now study the hypothesis — the "core + links" hypothesis — that no other complexes than these are formed. In the (p, q) diagram this means that all complexes that occur correspond to points on a straight line, passing through the composition of the core and with a slope depending on the composition of the link (see Fig. 1 c). In Fig. 1 d are given two specific examples: in one (black dots), the core is A_2B_2 and the link A_2B and thus the general composition of the complexes is $A_2B_2(A_2B)_n$. In the other (open circles), the core is AB and the link AB_2.

We shall characterize the line by its slope $1/t$, its intercept r with the p axis and its intercept s with the q axis (Fig. 1 c). These three quantities are related by

$$r = -ts \qquad (8)$$

and the equation of the line is

$$p = tq + r = t(q-s) \qquad (8\ a)$$

One may also say that the complexes can be written as $A_r(A_tB)_n$ or $B_s(A_tB)_n$, where r, s and t are constants and the number n stands for any number that will make p and q integers. Table 1 shows, for a few specific examples, how

Table 1. Formulas for complexes built up of "core" + n "links".
Possible ways of representing complexes

core	link	core(link)$_n$	$A_r(A_tB)_n$	$B_s(A_tB)_n$	r	s	t
A_2B_2	A_2B	$A_2B_2(A_2B)_n$	$A_{-2}(A_2B)_n$	$B(A_2B)_n$	-2	1	2
A_2B	AB	$A_2B(AB)_n$	$A(AB)_n$	$B_{-1}(AB)_n$	1	-1	1
A_3B_2	AB_2	$A_3B_2(AB_2)_n$	$A_2(A_{\frac{1}{2}}B)_n$	$B_{-4}(A_{\frac{1}{2}}B)_n$	2	-4	$\frac{1}{2}$
AB	AB_2	$AB(AB_2)_n$	$A_{\frac{1}{2}}(A_{\frac{1}{2}}B)_n$	$B_{-1}(A_{\frac{1}{2}}B)_n$	$\frac{1}{2}$	-1	$\frac{1}{2}$
B	A_tB	$B(A_tB)_n$	$A_{-t}(A_tB)_n$	$B(A_tB)_n$	$-t$	1	
A	A_tB	$A(A_tB)_n$	$A(A_tB)_n$	$B_{-1/t}(A_tB)_n$	1	$-1/t$	
—	A_tB	$(A_tB)_n$	$(A_tB)_n$	$(A_tB)_n$	0	0	

any "core + links" formula can be transformed to either one of these two forms. Of the constants, t is usually positive, whereas either r is negative, or s is negative, or both are zero. One or two of the constants or all may assume fractional values.

The "core + links" hypothesis does *not* require that *all* complexes on the line (8 a) are really formed in the solution. Perhaps only one or a few complexes really occur. What we have assumed is however that no complex A_pB_q is formed in appreciable amount which does not correspond to a point on the line (8 a):
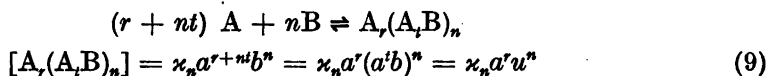
$$p = tq + r$$

We shall follow this hypothesis and see how the experimental data can be used for finding r and t. If the experimental data lead to and are consistent with a certain line $p = tq + r$, this should indicate that the "core + links" hypothesis, even if not absolutely true, is a useful approximation.

If only one or two polynuclear complexes exist, our hypothesis is automatically valid since it is always possible to draw a straight line through one or two points (see *e.g.* Fig. 1 e).

In passing, it may be noted that this description of the "core + links" complexes as $A_r(A_tB)_n$ or $B_s(A_tB)_n$ can be applied to mononuclear complexes, AB_n ($r = 1, t = 0$), but not so well to A_nB. This little lack of symmetry has no practical importance since a system with only mononuclear complexes is easily recognized (see below) and can be well treated by the simple methods already given in the literature.

Equilibrium equations

If the "core + links" hypothesis is valid, only those equilibrium constants $\varkappa_{pq}$ that correspond to points on the line (8 a) will differ from 0. We shall denote the equilibrium constant for the formation of $A_r(A_tB)_n$ by $\varkappa_n$.

$$(r + nt) \ A + nB \rightleftharpoons A_r(A_tB)_n$$

$$[A_r(A_tB)_n] = \varkappa_n a^{r+nt}b^n = \varkappa_n a^r(a^tb)^n = \varkappa_n a^r u^n \tag{9}$$

We have introduced for convenience the variable

$$u = a^t b \tag{10}$$

Introducing (9) into (4), we find

$$S = a^r \Sigma \varkappa_n u^n = a^r \varphi(u) \tag{11}$$

Equations (11) and (5), and (7) give

$$A - a = a \left(\frac{\partial S}{\partial a}\right)_b = a^r \left[r \ \varphi(u) + tu \ \varphi'(u)\right] = BZ \tag{12}$$

In the partial derivation of φ, (10) should be remembered. Similarly we find from (11) and (6)

$$B - b = b \left(\frac{\partial S}{\partial b}\right)_a = a^r u \varphi'(u) \tag{13}$$

Equations (12), (13), and (10) will be the basis for the following discussion and for our methods for finding r and t from the data.

TREATMENT OF DATA (a, A, B)

The $Z(\log a)_B$ curves; limiting values for Z

Let us assume that we have summarized our experimental data by plotting $Z = \dfrac{A-a}{B}$ as a function of $\log a$. We thus have a number of curves, on each of which the total concentration B has a constant value: B_1, B_2 etc (see Fig. 2).

As a matter of fact, in practically all systems hitherto studied these curves $Z(\log a)_B$ seem either to be all parallel with a spacing ($\varDelta \log a$) proportional to $\varDelta \log B$ or to coincide. These two types of behavior will be dealt with separately; they both logically lead to a "core + links" mechanism. Then some equations will be given which may be used for finding r and t in other cases, assuming the "core + links" hypothesis to be valid for them too.

Sometimes the concentration of one complex $A_r(A_tB)_n$, with one special value for n, may predominate over the concentrations of all other complexes and that of free B; then the average number of A bound per B is $Z = t + \dfrac{r}{n}$, as can be seen directly from the formula of the complex (and can also be derived from equations 11—13).

Sometimes the *concentration of free B is negligible* within the whole range studied — an example is the formation of aluminate complexes $OH(Al(OH)_3)_n^-$ ($r = 1$, $t = 3$, n may vary) [10,16] where the concentration of free Al^{3+} is negligible in the whole range in which these complexes have been studied. Then Z has the two limiting values (compare Fig. 2 b)

$$Z = \frac{A-a}{B} \nearrow\searrow \left. \begin{array}{l} t + \dfrac{r}{n_{\max}} \quad \text{for decreasing } \log a \\[2mm] t + \dfrac{r}{n_{\min}} \quad \text{for increasing } \log a \end{array} \right\} \begin{array}{l} \text{if } r > 0 \text{ and} \\ \text{always } b \ll B \end{array} \qquad (14)$$

Equation (14) may be derived from (11—13) assuming r to be positive, as it generally seems to be in such cases. (If r were negative, $n_{\max}$ and $n_{\min}$ would change places in (14).) If n can have any value between 1 and ∞, then the limits for Z are t and $t + r$.

If, on the other hand, *free B predominates* over all complexes *at one end* of the range studied, then (compare Fig. 2 a)

$$Z = \frac{A-a}{B} \nearrow\searrow \begin{array}{l} 0 \quad \text{for decreasing } \log a \ (b \approx B) \\[2mm] t + \dfrac{r}{n_{\max}} \quad \text{for increasing } \log a \ (b \ll B) \end{array} \qquad (15)$$

Here r is usually negative; if $n_{\max} = \infty$, the limits are 0 and t. (If r were to be positive, the limit for large a would be $Z = t + r/n_{\min}$.)

Parallel curves $Z(\log a)_B$

It is very often observed that the curves giving $Z(\log a)_B$ have the same shape, within the limits of experimental error, and are only shifted parallel with the $\log a$ axis. Moreover it is found that the spacing $\varDelta \log a$ between

two curves, say for B_1 and B_2, is proportional to the logarithm of the B ratio, $\Delta \log B = \log (B_1/B_2)$. Let us denote by R the constant ratio $\Delta \log B/\Delta \log a$. We then have

$$\left(\frac{\partial \log B}{\partial \log a}\right)_Z = R = \text{constant; } Z(a^{-R}B) \tag{16}$$

i.e. $Z = \dfrac{A-a}{B}$ must be a function of $a^{-R}B$.

Let us start with the general equations (5), (6), and (7), permitting *a priori* any formula A_pB_q. We find

$$Z = \frac{A-a}{B} = \frac{\Sigma p \varkappa_{pq} a^p b^{q-1}}{1 + \Sigma q \varkappa_{pq} a^p b^{q-1}} \tag{17}$$

$$a^{-R}B = a^{-R}(b + \Sigma q \varkappa_{pq} a^p b^q) = a^{-R}b(1 + \Sigma q \varkappa_{pq} a^p b^{q-1}) \tag{18}$$

If Z is to be a function of $a^{-R}B$ only, then the conditions are:

1) If b is *not negligible* in comparison with B, all terms $a^p b^{q-1}$ in the sums in (17) and (18) must be powers of $a^{-R}b$, which appears as a separate term in (18); thus $p = -R(q-1)$ or

$$p = -Rq + R \tag{19}$$

As seen by comparing (19) with (8 a) we would have "core + links" complexes with

$$R = r = -t; \ s = 1 \tag{19 a}$$

It is more convenient to write the formula of the complexes as $B(A_rB)_n$ than as $A_{-t}(A_rB)_n$. In this case the spacing gives both r and t. If only one complex A_PB_Q occurs, the $Z(\log a)_B$ curves also become parallel (eq. 29 below). This is best treated as a special case of $B(A_rB)_n$.

2) If b is *negligible* in comparison with B, the "1" disappears in the denominator of (17) and in the last term of (18). We may then multiply both sums in (17) by $a^{-R}b$ so that the sums in (17) and (18) contain the same terms $a^{p-R}b^q$. The condition for Z to be a function of $a^{-R}B$ is then that all these terms are powers of the same variable. We may always write this variable in the form $a^t b$, where t is a constant to be determined. We may thus from (16) deduce the equation

$$p-R = tq \tag{20}$$

By comparing (20) with (8 a) we find that we would still have "core + links" complexes; the general formula is $A_r(A_tB)_n$ where

$$r = R \tag{20 a}$$

may be found from the spacing of the curves. To find t one may use the limiting values of Z, and (14).

Consequently, if the family of curves $Z(\log a)_B$ turns out to be parallel, the chief product must be "core + links" complexes, and then the general formula can be obtained from the spacing and the limiting values. As a further step one should like to find out the details of the mechanism: whether one or several complexes are formed and the magnitude of the equilibrium constants.
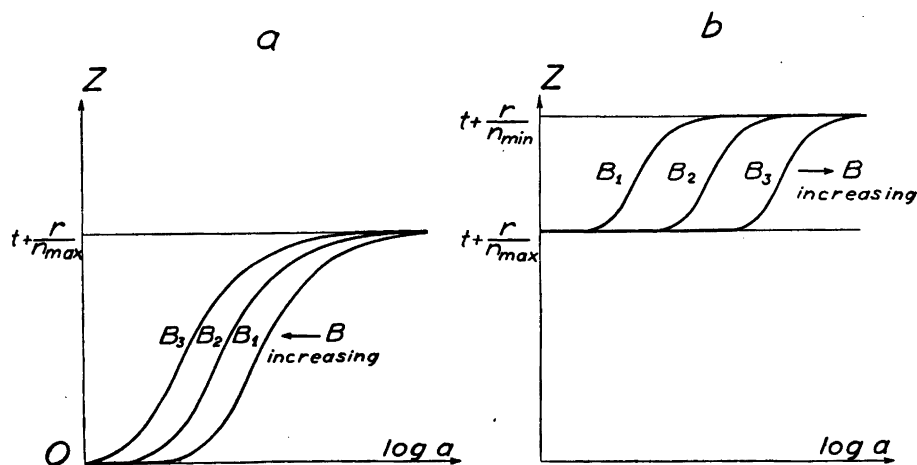
Fig. 2. Set of curves Z (log a)$_B$, schematical. For each curve, B has a constant value: B_1, B_2 etc. The horizontal spacing between two curves at a certain Z is denoted by $\triangle$ log a in the text. a) Parallel curves, corresponding to complexes $B(A_iB)_n$. b) Parallel curves corresponding to $A_r(A_iB)_n$ in a range where b is negligible.

A practical approach, since Z is a function of $a^{-r}B$ (16), is to plot some function y of Z versus some function x of $a^{-r}B$. Then all points should fall on a single curve $y(x)$ and our task would be confined to testing the experimental data against theoretical curves $y(x)$ calculated under various assumptions. It is advisable to eliminate the constants t and r from y and x so that the same theoretical curves y (x) can be applied to systems with different values for r and t. It is also desirable that the connection between y and x on one hand, and the measured quantities on the other, be so direct that one does not lose the feeling for the real accuracy and the possible interference of experimental errors. Suitable functions are given in (25) and (26).

1) *Complexes with B core, B(A,B)$_n$.* If $r = -t$, we may write the general formula for the complexes as $B(A_iB)_n$. Here t is a constant whereas n may vary. We then have, denoting by k_n the equilibrium constant for the formation of the n:th complex,

$$[B(A_iB)_n] = c_n = k_n b u^n \qquad (21)$$

$$u = a^i b \qquad (10)$$

and from (4 to 7), (21), and (10):

$$S = \Sigma c_n = b\Sigma k_n u^n = bf(u)$$

$$A - a = a\left(\frac{\partial S}{\partial a}\right)_b = ab \ \frac{df}{du}\left(\frac{\partial u}{\partial a}\right)_b = btuf' = BZ \qquad (22)$$

$$B - b = b(f + uf'); \quad B = b(1 + f + uf')$$

$$Z = \frac{A-a}{B} = \frac{tuf'}{1 + f + uf'} \qquad (23)$$

Here $f(u)$ is a function of u. If one compares the expressions for S in (11) and (22) and remembers (19 a) and (10), it is seen that $uf(u) = \varphi(u)$, the function in (11). It is easily verified that for Z constant, $u = a^t b = $ constant, $B = b \cdot$ constant, and thus $a^t B = $ constant, or

$$\left(\frac{\partial \log B}{\partial \log a}\right)_Z = -t \text{ in the whole range (for } r = -t) \tag{24}$$

This is equivalent to our previous equations (16) and (19 a).

A convenient way of plotting the experimental data is to give a diagram of $y(x)$ where (eq 22 and 23):

$$y = \frac{A-a}{Bt} = \frac{uf'}{1+f+uf'} \tag{25}$$

$$x = t \log a + \log B = \log u + \log (1+f+uf') \tag{26}$$

Since each of y and x is a function of u only, y must be a function of x only and all experimental points should be on the same curve $y(x)$.

This curve can then be tested against various assumptions concerning the equilibrium constants k_n. This will be the subject of part II in this series.[12]

1 a) *Only one complex* $A_P B_Q = B(A_t B)_N$. If only one complex is formed, say $A_P B_Q = B(A_t B)_N$, we have from (4) to (7)

$$\left. \begin{array}{l} S = [A_P B_Q] = \varkappa_{PQ} a^P b^Q \\ A - a = P \varkappa_{PQ} a^P b^Q = BZ \\ B - b = Q \varkappa_{PQ} a^P b^Q \end{array} \right\} \tag{27}$$

$$Z = P \; [Q + \varkappa_{PQ}^{-1} a^{-P} b^{1-Q}]^{-1} \tag{28}$$

Thus if Z is kept constant, $a^P b^{Q-1} = $ constant, $B - b = b \cdot$ constant, $B = b \cdot$ constant, $a^P B^{Q-1} = $ constant and

$$\left(\frac{\partial \log B}{\partial \log a}\right)_Z = \frac{-P}{Q-1} = -t \tag{29}$$

The curves $Z(\log a)_B$ will be parallel in the whole range. If $Q = 1$, they will even coincide. A single complex may be treated as a special case of 1), writing the complex as $B(A_t B)_N$ with $N = Q-1$, $t = P/(Q-1)$. Then in (21), all $k_n = 0$ except k_N which has a finite value. The same functions y and x as given in (25) and (26) may be used. To find out whether one or several complexes are formed one must analyse the curve $y(x)$, as will be discussed in Part II [12].

2) *Complexes* $A_r(A_t B)_n$; *b negligible*. Let us assume that b, the concentration of free B, is negligible in the whole range so that the only species present in measurable quantities are A and the complexes $A_r(A_t B)_n$. To get equations of the same form as in 1), we shall consider the "core" of the complexes to be the smallest possible complex, $A_{r+t} B = C$. We shall use

$[A_{r+t}B] = [C] = c;\quad [A_{r+t}BA_t B] = [CA_t C] = c_1;\quad \ldots\ldots$
$[A_{r+t}B(A_t B)_n] = [C(A_t C)_n] = c_n;\quad a^{-r} c = u$ \hfill (30)

We can write the reactions as reactions between A and C:

$$2A_{r+t}B \rightleftharpoons A_{r+t}BA_rB + rA \text{ or } 2C \rightleftharpoons CA_rC + rA$$

$$c_1 = k_1 c(a^{-r}c) = k_1 cu$$

$$(n + 1)A_{r+t}B \rightleftharpoons A_{r+t}B(A_rB)_n + rnA \text{ or } (n + 1) \ C \rightleftharpoons C(A_rC)_n + rnA$$

$$c_n = k_n cu^n \tag{31}$$

These equations are analogous to (21). In fact, our new $u = a^{-r}c$ differs only by an unknown constant factor (the equilibrium constant of $(r + t)A + B \rightleftharpoons C$) from $a^t b$.

The total amount of C present per liter, C, is equal to B. We shall let the subscript "c" indicate the quantities obtained if the reaction mixture is considered to be made up from A and C instead of from A and B. For instance, the total amount of A, A_c, is obtained by substracting from the previous A the amount bound in C. One can verify that

$$\left.\begin{aligned} A_c &= A - (r + t)C; \ C = B \\ S_c &= \overset{\infty}{\underset{1}{\Sigma}} c_n = c\Sigma k_n u^n = cf(u) \\ A_c - a &= a\left(\frac{\partial S_c}{\partial a}\right)_c = -rcuf' = Z_c C \\ C - c &= c\left(\frac{\partial S_c}{\partial c}\right)_a = c(f + uf') \end{aligned}\right\} \tag{32}$$

The two quantities y and x to be plotted are (eq. 32)

$$y = \frac{A_c - a}{C \cdot (-r)} = \frac{r + t - \dfrac{A-a}{B}}{r} = \frac{uf'}{1 + f + uf'} \tag{33}$$

$$x = -r \log a + \log C = -r \log a + \log B = \log u + \log(1 + f + uf') \tag{34}$$

Equations (33) and (34) have the same form as (25) and (26). The same theoretical curves $y(x)$ as for the $B(A_rB)_n$ complexes can thus be used for finding the function $f(u)$ for the $C(A_rC)_n$.

Coinciding curves $Z(\log a)_B$

Suppose that the curves giving $Z = \dfrac{A-a}{B}$ as a function of $\log a$ are not only parallel for various B, but also actually coincide. This means that Z is a function of a only. Since the partial derivative R in (16) becomes infinite, the formula (19) for the complexes becomes indefinite. We shall return to the general equations (5), (6), and (7), admitting a priori any complex A_pB_q

$$Z = \frac{A-a}{B} = \frac{\Sigma p \varkappa_{pq} a^p b^q}{b + \Sigma q \varkappa_{pq} a^p b^q} \tag{17 a}$$

What are the conditions that Z is to be a function of a only?

1) If b *is not negligible,* it is necessary that the sums are of first order with respect to b, so that b can be eliminated. This means that q is always $= 1$, so that the complexes are only *mononuclear*, A_nB.

2) If b *is negligible* in the denominator in the whole range, b can be eliminated from (17 a) provided only one value for q exists. This value need not necessarily be 1. Thus the complexes must be *"homonuclear"*, either only A_nB or only A_nB_2 etc. (If $p/q = $ constant, $Z = $ constant. In this trivial case all Z curves coincide in a horisontal line — it is met with in a concentration range where b is negligible, and either one complex predominates, or all complexes have the formula $(A_tB)_n$.)

The methods for dealing with mononuclear complexes have been treated so fully by other authors that there is no need to go into them here. The adaption of these methods to homonuclear complexes is easy, if the need should arise.

General equations for finding r and t from (a, A, B) data

At present it may not be profitable to discuss in detail all other possible ways in which the family of curves $Z(\log a)_B$ might behave. We shall assume, as a working hypothesis, that the "core + links" hypothesis is valid. We shall then limit ourselves to some general relationships which may be helpful for finding r and t from any set of curves and also discuss how to recognize a set of curves with $r = 0$ (general formula $(A_tB)_n$). Should the need arise, the general equations given here can certainly be applied to other special cases as well.

The formulas for the *limiting values for Z,* given in (14) and (15), may also be useful for finding r and t.

Spacing for $b \ll B$. If, in a certain range of the curves, the concentration of free B, b, can be neglected in comparison with the total concentration B, we find from (12) and (13) $Z = t + r \dfrac{\varphi(u)}{u\varphi'(u)}$. If Z is not invariable (see below) it must be a function of u only. Consequently, if Z is kept constant, u is constant, and from (13) neglecting b: $B = a^r u\varphi'(u) = a^r \cdot$ constant, or $\log B = r \log a + $ constant. This result may be written in differential form:

$$\left(\frac{\partial \log B}{\partial \log a}\right)_Z = r \text{ for } b \ll B \tag{35}$$

and is identical with equations (16) + (20 a).

In general one may expect (35) to be approached with increasing Z. However, if the ratio p/q is the same in all complexes, then b is not negligible in comparison with B until Z becomes constant. This complication will occur either if only one complex is formed, say A_PB_Q, or if $r = 0$ and thus all complexes have general formula $(A_tB)_n$. Each of these cases is dealt with separately in this paper.

Spacing if B and one complex predominate. If we can neglect all molecular species except A, B, and one complex, say $A_P B_Q = A_r(A_tB)_N$, then we have already proved (see 29 where however we set $r = -t$):

$$\left(\frac{\partial \log B}{\partial \log a}\right)_z = -\frac{P}{Q-1} = -\frac{r + Nt}{N-1} \tag{29 a}$$

This equation is, of course, useful only when b is not negligible in comparison with B; otherwise Z is constant $= P/Q$, giving a horizontal line.

Spacing for $b \approx B$. In the region where complex formation just starts, and thus Z is small, we may assume that practically only B and the lowest complex, with $n = n_{min}$, exist. Then from (29 a)

$$\left(\frac{\partial \log B}{\partial \log a}\right)_z \rightarrow -\frac{r + n_{min}t}{n_{min}-1} \text{ for } b \approx B \tag{36}$$

If $n_{min} = 1$, one of two cases may occur: (a) $r = -t$, which makes the expression (36) equal to $-t$. This case, which gives a series of parallel curves, has been dealt with. (β) $r + t \neq 0$, which means that the derivative in (36) is infinite and that the spacing between the curves along the log a axis approaches zero for decreasing Z.

Curves for $a^{-r}(A-a)$. If one wants to test a certain value for r and to find t, one may calculate the quantity $a^{-r}(A-a)$ and plot it as a function of log a with one curve for each B value. The following relationship may prove a useful test:

$$\left(\frac{\partial \log B}{\partial \log a}\right)_{a^{-r}(A-a)} \begin{array}{l} \nearrow r \text{ for } b \ll B \\ \searrow -t \text{ for } b \approx B \end{array} \tag{37}$$

As seen from (12), $a^{-r}(A-a)$ should be a function of u only. When $a^{-r}(A-a)$ is kept constant, $u = a^tb$ is constant and $B-b = a^r \cdot$ constant. For $B \gg b$, $B = a^r \cdot$ constant, which is the first relationship in (37). For $b \approx B$, on the other hand, $u \approx a^tB =$ constant, which gives the second relationship.

Complexes $(A_tB)_n$: $r = 0$. If the general formula of the complexes is $(A_tB)_n$, $r = 0$ and we have from (12) and (13)

$$\begin{array}{l} A-a = tu\varphi'(u) = tBy \\ B-b = u\varphi'(u) = By \end{array} \tag{38}$$

where $y = Z/t$ as in (25). We see that $b = B(1-y)$; thus b cannot be neglected in comparison with B until y becomes 1 and $Z = t$. In this special case we have no help of (35) in determining r and t.

However, if it is suspected that $(A_tB)_n$ is the general formula, one can test it, e.g., in the following way. The value for t may be obtained either from (37) or as the upper limit of Z. Then we can calculate for each point $b = B-\frac{A-a}{t}$. Now, if $t \log a + \log\left(B-\frac{A-a}{t}\right) = t \log a + \log b = \log u$ is plotted on the horizontal axis and $A-a$ (or log $(A-a)$) on the vertical, all experimental points should fall on one curve since $A-a$ should be a function of u only.

TREATMENT OF DATA (a,b,A,B)

If both a and b can be measured very accurately, and if the "core + links" hypothesis is assumed to be valid, one may devise several ways for finding r and t based on equations (11 to 13). We shall consider a rather unfavorable case: a is measured, but the difference $(A-a)$ cannot be calculated with enough accuracy to be of much help. Our main source of information will be a set of curves giving, for a series of constant B values, $\frac{b}{B}$ (or better $\log \frac{B}{b}$) as a function of $\log a$.

For low a values, both of these functions give horizontal lines, $\frac{b}{B} = 1$, and $\log \frac{B}{b} = 0$. With increasing $\log a$, $\frac{b}{B} \to 0$ and $\log \frac{B}{b} \to \infty$. At very high values for $\log a$, $b \ll B$ and the complex $A_r(A_tB)_n$ with maximum value for n, n_{max}, predominates provided r is negative, as usual; if $r > 0$, n_{min} predominates. Then we find, e.g., from (27)

$$\left(\frac{\partial \log \frac{B}{b}}{\partial \log a}\right)_B = -\left(\frac{\partial \log b}{\partial \log a}\right)_B = \frac{P}{Q} = t + \frac{r}{n_{max}} \tag{39}$$

Thus if n is unbounded, the maximum slope is t.

Parallel curves $\log \frac{B}{b}$ $(\log a)_B$.

Let us assume that the curves $\log \frac{B}{b}$ $(\log a)_B$ for different B values are parallel with spacing $\varDelta \log a$ proportional to $\varDelta \log B$. The curves of $\frac{b}{B}$ $(\log a)$ or any other function of $\frac{B}{b}$ are then also parallel, with the same spacing. This may be expressed as

$$\left(\frac{\partial \log B}{\partial \log a}\right)_{\frac{B}{b}} = -T = \text{constant or } \frac{B}{b} \ (a^T B) \tag{40}$$

We shall start with (6), assuming a priori that all complexes A_pB_q may exist. We find

$$\frac{B}{b} = 1 + \Sigma q \varkappa_{pq} a^p b^{q-1}; \quad a^T B = a^T b + \Sigma q \varkappa_{pq} a^{p+T} b^q \tag{41}$$

The condition for $\frac{B}{b}$ to be a function of $a^T B$ only is that all terms $a^p b^{q-1}$ are powers of $a^T b$. Even if $b \ll B$, no new possibilities arise in this case. Thus

$$p = T(q-1) \text{ for all complexes} \tag{42}$$

By comparing (42) with (8 a) we find that we have "core + links" complexes with

$$r = -T, \; s = 1, \; t = T \tag{42 a}$$

The general formula for the complexes is then $B(A_rB)_n$ with $t = T$. This is the same as case 1 which we met with for parallel $Z(\log a)_B$ curves. We can apply equations (21) and (22) with $r = -t$. Using (22) we verify (the proof is similar to that of (24)):

$$\left(\frac{\partial \log B}{\partial \log a}\right)_{\frac{B}{b}} = -t \tag{43}$$

We can reduce our data to a single curve, for further discussion, by plotting

$$\eta = \log \frac{B}{b} = \log(1 + f + uf') \tag{44}$$

as a function of

$$x = t \log a + \log B = \log u + \log(1 + f + uf') \tag{26}$$

or of $\log u = t \log a + \log b$.

The analysis of such curves will be treated in part II [12].

The case with *only one complex* A_PB_Q is conveniently treated as a special case of $B(A_rB)_n$. With (27) or (43) we find

$$\left(\frac{\partial \log B}{\partial \log a}\right)_{\frac{B}{b}} = -t = -\frac{P}{Q-1} \tag{45}$$

Coinciding curves $\log \frac{B}{b} (\log a)_B$

If the curves $\log \frac{B}{b}(\log a)$ for various B coincide strictly, then $\frac{B}{b}$ is a function of a only. From (41) we find that this is possible only if always $q = 1$ which means that all complexes must be *mononuclear* in B: A_nB.

General equations for finding r and t

The *limiting slope* of the $\log \frac{B}{b}(\log a)_B$ curves, according to (39), gives $t + rn_{max}^{-1}$ if $r < 0$; $t + rn_{min}^{-1}$ if $r > 0$.

Spacing for $b \approx B$ and $b \ll B$. In analogy with (3 6) we find from (13) and (11) for negative r; for positive r, "min" and "max" change place:

$$\left(\frac{\partial \log B}{\partial \log a}\right)_{\frac{B}{b}} \nearrow\searrow \begin{array}{l} -\dfrac{r + n_{min}t}{n_{min}-1} \text{ for } b \approx B \\[2mm] -\dfrac{r + n_{max}t}{n_{max}-1} \text{ for } b \ll B \end{array} \tag{46}$$

If $n_{min} = 1$, the lateral spacing between the curves disappears for $b \approx B$, unless $r = -t$, in which case the curves are always parallel as discussed earlier. If $n_{max} = \infty$, the limiting spacing, like the limiting slope, will give t.

Curves for $b^{-s}(B-b)$. Using a tentative value of $s = -r/t$, one may plot $b^{-s}(B-b)$ as a function of $\log a$ to give a family of curves, either with B or b constant. Since B is usually kept constant in each experimental series, the curves with B constant are obtained immediately, whereas those with b constant must be calculated by interpolation.

As one may derive from equations (13) and (8), $b^{-s}(B-b)$ should be a function of u only, if one has hit on the correct s value. One may use the equations

$$\left(\frac{\partial \log B}{\partial \log a}\right)_{b^{-s}(B-b)} \quad \begin{matrix} \nearrow & -t \text{ for } b \approx B \\ \searrow & -ts = r \text{ for } b \ll B \end{matrix} \tag{47}$$

$$\left(\frac{\partial \log b}{\partial \log a}\right)_{b^{-s}(B-b)} = -t \text{ in the whole range} \tag{48}$$

One may then find out whether the values for t and r obtained from (47) and (48) with a tentative s are compatible with each other and with the s assumed.

Curves for $\dfrac{B-b}{A}$. With the aid of data (a, b, A, B) one may plot $\dfrac{B-b}{A}$ as a function of $\log b$, for a series of values $A_1, A_2 \ldots \ldots$, and study the limiting values and the spacing. The formulas are analogous to (14 to 36) except that B and A are interchanged and the general formula of the complex $A_r(A_tB)_n$ is replaced by $B_{r'}(AB_{t'})_{n'}$, where $r' = s = -r/t$, $t' = 1/t$, and $n' = r + nt$. The transformation of (35) and (36) gives, for example

$$\left(\frac{\partial \log A}{\partial \log b}\right)_{\frac{B-b}{A}} \begin{matrix} \nearrow & r' = -r/t \text{ for } a \ll A \\ \searrow & \dfrac{r' + n'_{min}t'}{n'_{min}-1} = -\dfrac{n_{min}}{r + n_{min}t-1} \text{ for } a \approx A \end{matrix} \tag{49}$$

If the measurements have been made for a series of constant B values, one must interpolate to get curves with constant A.

Curve for $a^{-r}(B-b)$. If one wants to test whether a certain set (r, t) is consistent with the data, one may calculate and plot $a^{-r}(B-b) = u\varphi'(u)$ versus $t \log a + \log b = \log u$. All experimental points should then be on a single curve, irrespective of B.

SUMMARY

A system is discussed in which two reactants A and B form a series of polynuclear complexes A_pB_q. The mathematical relationship between the total concentrations (A, B) and the free concentrations (a, b) of the reactants is given. A useful function called the complexity sum, S (equation 2) is introduced.

If only a, A, {and B are known, one may draw curves giving $Z = \dfrac{A-a}{B}$

as a function of $\log a$, with B constant for each curve. It is often observed that

these curves, with good approximation, are all parallel with spacing $\Delta \log a$ proportional to $\Delta \log B$. This behavior leads to the following conclusions: 1) If b is not negligible in comparison with B, all complexes present in appreciable amounts must have the general formula $B(A_tB)_n$; here t is a constant, which is obtained from the spacing of the curves (eq 24), whereas n may vary. 2) If the curves are parallel and b negligible in the whole range, the general formula must be $A_r(A_tB)_n$, where t and r are constants, and n variable; r is obtained from the spacing.

If the $Z(\log a)_B$ curves coincide and b is negligible, all complexes must be homonuclear, thus having the same q. If b is not negligible, they must moreover all be mononuclear in B, of general formula A_nB.

Equations are given for the case where all complexes can be written as built up from a certain "core" and a varying number of identical "links" A_tB. All existing complexes might then be written as $A_r(A_tB)_n$ or $B_s(A_tB)_n$, with $r = -ts$, but it is not necessary that all n values really occur. It is shown how the values for r and t can be found from measurements of either (a, A, B) or (a, b, A, B) in a specific system.

If the curves giving $\log \dfrac{B}{b}(\log a)_B$ are parallel, the formula for the complexes must be $B(A_tB)_n$ with t constant, n variable. If, moreover, they coincide, then only mononuclear complexes A_nB are possible.

This paper is limited to finding the general formula for the complexes, whilst the detailed investigation of the mechanism and the equilibrium constants *etc.* will be discussed in subsequent papers.

REFERENCES

1. Bjerrum, N. *Kgl. Danske Videnskab. Selskabs Skrifter, Naturvidenskab. math. Afdel.* (7) **12** (1914—1915) No. 4.
2. Bjerrum, J. *Diss.* Copenhagen 1941.
3. Leden, I. *Diss.* Lund (1943).
4. Bjerrum, N. *Diss.* Copenhagen 1908.
5. Bjerrum, N. and Faurholt, C. *Z. physik. Chem.* **130** (1927) 584.
6. Fronaeus, S. *Diss.* Lund 1948.
7. Ahrland, S. *Acta Chem. Scand.* **3** (1949) 374.
8. Neuss, J. D. and Rieman, W. III, *J. Am. Chem. Soc.* **56** (1934) 2238.
9. Granér, F. and Sillén, L. G. *Acta Chem. Scand.* **1** (1947) 631.
10. Brosset, C. *Acta Chem. Scand.* **6** (1952) 910.
11. Leden, I. *To be published.*
12. Sillén, L. G. *Acta Chem. Scand.* **8** (1954) 318 (Part II).
13. Hietanen, S. *To be published.*
14. Rydberg, J. *Acta Chem. Scand.* **4** (1950) 1503.
15. Fronaeus, S. *Acta Chem. Scand.* **5** (1951) 859.
16. Brosset, C., Hietanen, S. and Sillén, L. G. *To be published.*