(reported [12] 1 790 cm^{-1} and 1 850 cm^{-1}). The addition compound between triphenyl phosphine and maleic anhydride has two strong bands at 1 702 cm^{-1} and 1 805 cm^{-1}. We assume that these two bands are due to the coupling between two carbonyl groups and suggest therefore that the addition compound is a phosphorus ylid namely compound No. 12 in Table 1. This structure may also explain the unreactivity of 1,2-substituted maleic anhydrides [9].

1. Pinck, L. A. and Hilbert, G. E. *J. Am. Chem. Soc.* **69** (1947) 723.
2. Ramirez, F. and Dershowitz, S. *J. Org. Chem.* **22** (1957) 41.
3. Ford, Jr., J. A. *Org. Chem. Bull.* **31** (1959) No. 1.
4. Witting, G. and Schöllkopf, U. *Chem. Ber.* **87** (1954) 1318.
5. Wilson, B. D. *Org. Chem. Bull.* **32** (1960) No. 2.
6. Michaelis, A. and Köhler, E. *Ber.* **32** (1899) 1566.
7. Michaelis, A. and Gimborn, H. v. *Ber.* **27** (1894) 272.
8. Aksnes, G. *Acta Chem. Scand.* **15** (1961) 438.
9. Schönberg, A. and Ismail, A. F. A. *J. Chem. Soc.* **1940** 1374.
10. Ramirez, F. and Dershowitz, S. *J. Am. Chem. Soc.* **78** (1956) 5614.
11. Hartwell, E. J., Richards, R. E. and Thompson, H. W. *J. Chem. Soc.* **1948** 1436.
12. Norman Jones, R. and Sandorfy, C. in West, W. *Technique of Organic Chemistry* Volume IV p. 456. Interscience, New York 1956.
13. Davison, W. H. T. *J. Chem. Soc.* **1951** 2456.
14. Schönberg, A. and Michaelis, R. *Ber.* **69** (1936) 1080.

"Pit-mapping" — a General Approach for Computer Refining of Equilibrium Constants

DAVID DYRSSEN, NILS INGRI and LARS GUNNAR SILLÉN

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

The problem of determining the equilibrium constants for a number of simultaneous equilibria in solution may be expressed as a special case of the following well-known general problem. Having measured a quantity y and assumed a functional relationship:

$$y = f(k_1, k_2..k_r..k_N; a_1, a_2...) \qquad (1)$$

between y and some quantities $a_1, a_2 \ldots$ that can be varied but whose values are assumed to be known accurately in each experiment, one seeks to determine the set of unknown constants, $k_1 \ldots k_N$.

Provided a) that the expression for f is correct, b) that there are no other errors (especially no systematic ones) except the random errors in y, c) that the errors in y are "normally" distributed around the correct value, and d) that the weight w_i given to each measurement y_i is inversely proportional to the square of its standard deviation; then the "best" values for $k_1 \ldots k_N$ would be those that minimize the error-square sum:

$$U = \Sigma_i w_i (y_i - f(k_1 \ldots k_N; a_{1i}, a_{2i} \ldots))^2 \qquad (2)$$

Even though one can seldom hope that all these assumptions are correct, the "least-squares" condition, $U = $ minimum, is often employed in calculations. It is "objective" in the sense that the same set of data, and the same assumptions, should give the same answer in the hands of different workers. Perhaps, however, one should not forget the risk of over-estimating the significance of the answer.

If f is a linear function of the k's, the "least-squares" condition leads to linear equations, which can be solved by straightforward methods. For instance, in studies of equilibria of step-wise complex formation, if one has measured the concentration of the central atom, or of one specific complex (as it is, for instance, usually done in extraction studies), this condition leads to

linear equations which can be solved by means of the "least-squares" programs of various high-speed computers [1-3].

If the function f is not linear, the general trend has been to reduce it to a linear one, either by transformation (if possible), or, using the Gauss method, by expressing the derivatives of U by means of Taylor series, and using only the first terms. Programming a computer for the Gauss method is not always simple, and its performance is capricious [4].

In systems with polynuclear complexes between two reagents A and B, one usually meets with the two equations

$$B = b + \Sigma q \beta_{pq} a^p b^q; \ BZ = \Sigma p \beta_{pq} a^p b^q \quad (3)$$

In (3), B is the total concentration of B, a and b are the concentrations of free A and B, p and q are integers, β_{pq} are a set of equilibrium constants, and Z is the average number of A bound per B. The data are usually of the form $Z(B, a)$, so that B and a are known rather accurately. Since b must be eliminated from the equations, it seems hard to attack problems like this by straight-forward "least-squares" equations.

However, one can make oneself free from the condition that the equations should be linear, if one considers directly the error-square sum U in (2) and studies the function

$$U(k_1, k_2 \ldots k_N) \quad (4)$$

This is done in a family of computer programs "LETAGROP" which we are now developing, using the Ferranti-Mercury autocode.

The principle is as follows; it is so simple that it may have been used earlier by others without our being aware of it.

By means of graphical methods — which we consider as the most important part of the data treatment — we conclude which complexes are present or may be suspected to exist in appreciable amounts, and obtain a set of approximate equilibrium constants.

The computer then systematically varies each constant in turn, keeping the others constant, until it finds the value that gives the lowest value for U. Then it passes to another constant and etc; it can be made to return to a certain constant and re-adjust it later on. The procedure gives not only the lowest point:

$$\mid U(\min = U_0 = U(k_1'', k_2'' \ldots k_N'') \quad (5)$$

but also a map of the surroundings.

For linear expressions f in (1), in particular, $U(k_1 \ldots k_N)$ is a second-degree surface: if $N = 1$, it is a parabola, with $N = 2$ an elliptic paraboloid, for higher N it is a generalized elliptic paraboloid in $(N + 1)$-dimensional space. We shall for any value of N refer to the lowest point U_0 and its immediate surroundings as the "pit":

A suitable measure of the spread of the data is the range around U_0 enclosed by what we may call the "D boundary"

$$U = U_0 + U_0(n-N)^{-1} \approx U_0 + U_0/n \quad (6)$$

Here, U_0 is the value of U at the lowest point (5) of the pit, and n is the number of experiments. In our experiments, as a rule, $n \gg N$. The projection of the D boundary along the U axis is two points if $N = 1$, an ellipse if $N = 2$, an ellipsoid if $N = 3$, and for higher N a generalized ellipsoid in N dimensional space $(k_1 \ldots k_N)$, with the "least-squares point" $(k_1'' \ldots k_N'')$ as its centre.

The range of values that each k_r can assume on the D boundary is identical with the range of its "standard deviation" $k_r'' \pm D[k_r]$, whatever connection this may have with chemical reality.

For non-linear cases it seems a permissible approximation to assume the "pit" to be a second-degree surface in the range close to U_0. By proper programming, the computer can be made to search for the lowest point of the "pit" within any reasonable precision; assuming the U surface to be of the second degree, one can calculate directly the position of the lowest point, and the extension of the D boundary, given a certain minimum number of U values (three for $N = 1$, six for $N = 2$, etc).[*]

If a certain k_r exerts an influence only in a limited region of the data field, the program allows one to screen out the other points, and then to consider the contribution to U from that part only.

The same main programs "LETAGROP" can be used for various non-linear problems. The special equations valid for each problem need only be stated in a short additional program and are then solved by the general method of our program "KUSKA" which forms a loop in LETAGROP. So, even though the machine may have to work several minutes more than in a straight-forward solution, great gains are made in the human programmer's time.

In the future we intend to apply these computer programs to a number of various chemical problems, especially the following:

---

[*] Note added in proof. This procedure has proved much faster than adjusting the constants one by one.

a) The refinement of the equilibrium constants in systems with polynuclear complexes, where approximate values have been found graphically.

b) Calculation of consecutive equilibrium constants in systems with mononuclear complexes, where only the ligand number has been determined: even if linear transformations can be made [3b] the pit-mapping method seems a worth-while alternative. Even spectrophotometric data, which commonly give non-linear expressions may be attacked by this method.

c) Calculation of the parameters in the "two parameters" approximation [5] for equilibria, where the available data are not accurate enough for determining, independently, the formation constants for all the complexes MA, MA_2, MA_3 ... MA_N. For such systems, two ways of approximate calculations have been proposed:

1) When the standard deviation of an individual equilibrium constant comes out as greater than the value of this constant, Sullivan, Rydberg and Miller [3b] eliminate that particular constant from their equations, which from a chemical point of view is equivalent to assuming that the concentration of the related species is equal to zero.

2) In the "two parameters" approximation [5], one assumes that all complexes exist and that the ratio between consecutive formation constants is a constant. — Neither of the approximations 1) or 2) can be completely true, but the second one may, we think, often come somewhat closer to chemical reality. Earlier, only the first approximation could be treated by computer programs [3].

Our results hitherto like, in our opinion, those of Sullivan et al.[3b] indicate that with good data, the "least-squares" approach gives no gain in real accuracy, as compared with adequate graphical methods, using the experimental data directly. In problems with many constants to be determined, considerable time may however be gained by making the final refinement of the constants by means of the computer.

The results and the full description of our program, will be published in due time.

1. Panthaleon van Eck, C. *Thesis*, Amsterdam 1958.

2. McMaster, D. L. and Schaap, W. B. *Proc. Indiana Acad. Sci.* **67** (1958) 111.

3a. Rydberg, J. and Sullivan, J. C. *Acta Chem. Scand.* **13** (1959) 186.

3b. Sullivan, J. C., Rydberg, J. and Miller, W. F. *Acta Chem. Scand.* **13** (1959) 2023.

3c. Rydberg, J. and Sullivan, J. C. *Acta Chem. Scand.* **13** (1959) 2057.

3d. Rydberg, J. *Acta Chem. Scand.* **14** (1960) 157.

4. Moore, R. H. and Ziegler, R. K. *LA-2367*, 1960.

5. Dyrssen, D. and Sillén, L. G. *Acta Chem. Scand.* **7** (1953) 663.

Investigation of the Structure of Trifolirhizin, an Antifungal Compound from *Trifolium pratense* L.

J. B-SON BREDENBERG and
PENTTI K. HIETALA

Laboratory of the Foundation for Chemical Research, Biochemical Institute, Helsinki, Finland

In an investigation of the antifungal substances in red clover (*Trifolium pratense* L.) the isolation of one of the main components, $C_{21}H_{24}O_{10}$, m.p. 139—140°, has been reported [1]. This compound has now been investigated more closely. Since it is apparently a new compound, isolated so far only from the roots, the name trifolirhizin is proposed for it.

Trifolirhizin crystallizes from methanol as rods, m.p. 142—144° (decomp.), $[a]_D^{20}$ —183°, with one mole of methanol. The elementary composition of the compound has now been found to agree with $C_{22}H_{22}O_{10}$. It does not contain methoxyl or C-methyl groups and, according to the chemical and infrared data, is free from carbonyl and carboxyl groups. Upon catalytic hydrogenation it is recovered unchanged. It forms a tetraacetate, $C_{30}H_{30}O_{14}$, m.p. 188—189,° $[a]_D^{20}$ —126° in which all the hydroxyl groups are acetylated. Upon acid hydrolysis it forms glucose and ether-soluble substances which could not be characterized. Alkaline hydrolysis probably also releases glucose. Trifolirhizin is thus a glucoside.

Taking the glucose molecule into account, the empirical formula for the agly-