AIMS, APPLICATIONS AND ACHIEVEMENTS. AN INTRODUCTORY ESSAY

Jens G. REICH
Berlin, G.D.R.

1. Introduction

To start with I must say that the FEBS Summer School on Computing Techniques in Biochemistry was a complete success. Students with small experience in the field profited greatly from detailed technical instruction as well as from plenary lectures, whereas the "professionals" had extensive opportunities to discuss their problems and results. The presence of beginners and specialists at the same summer school will have helped to stop the development of a "knowledge gap" between them, which could become very harmful in view of the expanding applications of computing techniques to Biochemistry. The organisers, headed by J.H.Ottaway, had much more to do than for the usual run of meetings of this size, since, apart from the regular arrangements, a complicated system of interconnected plenary lectures, specialized seminars and computer instructions had to be prepared and run. Everybody with computer experience can appreciate the efforts of the organisers, from the Departments of Biochemistry, Animal Genetics, and Computer Science of the University of Edinburgh, to "debug" and test the programs. And, last but not least, the organizers succeeded in assembling a teaching staff of most distinguished specialists from the United Kingdom, U.S.A., and West Germany.

The participants included large groups from the U.K. and West Germany together with smaller numbers from other European countries. One of the main problems was communication: in fact the marriage between Computer Science and Biochemistry is by no means easy. This is not due to lack of personal relations but rather to the differences of background knowledge and nomenclature. But it could be seen at the summer school that this gap can be filled by efforts on both sides. In fact, the best lectures did not lead a discouraged audience through a mass of facts allegedly mandatory for doing the job, but rather gave a simple, fully referenced survey of the applicable methods, without understating the problems but encouraging attempts at their solution.

Thirteen lectures and a large number of discussions and free communications provided a complete survey of the techniques in use. Quantum biochemistry and computational techniques in protein crystallography were omitted, a wise restriction in my opinion, preventing a hopeless extension of the field. All lectures and discussions were admirably run by S.Michaelson as mathematician and by H.Kacser as biologist.

To give a brief review of the application of computing techniques as presented at this meeting is not easy. I had the impression of a difference which exists between a theoretical approach to the general properties of computer-simulated biochemical systems on the one hand, and pure practical computing to extract parameters from a set of data on the other. The two lines should not diverge too far, since the former tends to art for art's sake, and the latter to overlook hidden theoretical inconsistencies in the numerical data. A good example of the latter danger is that sigmoidal enzyme kinetics, so important nowadays, have been overlooked for decades, because wishful numerical thinking led people to suppress deviations from linearity in reciprocal plots.

B.Hess pointed out that one should keep in mind the time scale in question. Problems in "millisecond-biochemistry" such as the physicochemical description of enzyme mechanisms cannot be investigated at the same time as "second-biochemistry" (regulation of pathway throughput, stationary fluxes, oscillations) or, *a fortiori*, "long-term-biochemistry" such as enzyme adaptations, etc. This principle of classification is retained in the following description.

2. Survey of currently used computer methods

2.1. *Synthesis and analysis of enzyme mechanisms*

E.M.Chance presented and made available a large-scale program designed to simulate enzyme mechanisms in the millisecond range. The program is a generalisation of the well-known Garfinkel-Hess model of the glycolytic pathway [1], which was also described in Edinburgh by one of its originators, B.Hess, who covered the biochemical problem of feeding the computer with appropriate data. The programs of these authors are developed to a remarkable degree of abstraction which, paradoxically, makes possible a conversation between experimenter and computer in a concrete, problem-oriented language which is easily learned without mathematical training [2–4]. A specified mechanism and a set of initial data form the input, and the computer replies with graphs, or lists, of component concentrations versus time, which can be compared with the data. A frequent technical problem is the appearance of "stiff" systems which result in very slow progress to the solution (see below).

2.2. *Synthesis and analysis of intermediary pathways and cycles*

In the second range of characteristic times, the enzyme intermediates are usually assumed to be in the Briggs-Haldane steady-state. This greatly simplifies the mathematical treatment, permitting large systems to be studied. J.Burns and H.Kacser described a program to simulate intermediary pathways and cycles which may be used to calculate the stationary levels of metabolites and steady-state output and throughput fluxes, given a set of input data (fluxes and enzyme activities). Alternatively, the latter may be calculated, given the levels of metabolites and the net flux rate. The input language is remarkably simple. The experimenter can easily study the effects of isoenzymes, of different K_m values, of mutant deficiencies, of altered milieu conditions, and much more. An ingenious procedure is used in this program, which circumvents the involved physicochemical theory (Monod's model etc.) that would otherwise be required to describe effectors. The user specifies the characteristics observed (say, for an inhibition effect, the percentage of maximal inhibition, the necessary concentration for half-maximal effect , and the steepness and sigmoidicity of the kinetic curve) and the computer formulates an operational expression, approximating the actual kinetics, and calculates the behaviour of the whole system – behavior which can scarcely be apprehended on a purely intuitive basis. Such studies seem to be of importance in view of the numerous reports of more or less adventitious distant effects of metabolites on enzymes which are claimed to exercise some control on the pathway involved. The Edinburgh group adopts a sensitive coefficient

$$C_{M_j}^{F_i} \equiv \frac{\delta F_i}{\delta M_j}$$

for the dependence of a flux F_i on a metabolite M_j, or an analogous coefficient $C_{E_k}^{F_i}$ on an enzyme E_k, to study such effects quantitatively. For example, if a 5% increase in M_j would produce a 1% decrease of F_i, then $C_{M_j}^{F_i} = -0.01/0.05 = -0.2$. The computer calculates a matrix of such coefficients by adding small increments to the parameters. The procedure is approximate, since C depends on the step size in the non-linear case, but the heuristic value of this formalism for the quantitative understanding of such effects is obvious.

2.3. *Synthesis and analysis of compartmental systems*

M.Berman has developed a fascinating program for extraction of parameters as well as for prediction of the behaviour and assessment of the validity of a mathematical model fitted to the data. The subtleties of this large-scale program need not be considered, since Berman himself has published a detailed survey [5]. Though the program can process nearly all reasonable types of data, it seems to me to be especially suited to cover the longer time-scale, such as inter-organ exchanges of metabolites under the influence of hormones and conditions in the milieu (examples are the kinetics of iron, of calcium, and of iodine, glucose turnover, metabolism of fatty acids and other lipids), rather than studies within single organs or cells. There are close relationships to physiological problems such as distribution kinetics in the water compartments of the body and circulation kinetics of dyes. The dynamic system is usually formulated by linear differential equations but algebraic and other

models may equally be solved. There is a large subprogram of statistical operations which allows a detailed judgment of the consistency and uniqueness of the model applied to be made. A disadvantage is that rather exact preliminary estimates of the rate constants are required to guarantee convergence of the optimization process.

It should be mentioned that simpler but also less versatile programs may be used for compartmental analysis, and that analog computers may also be useful [6] especially for linear models (e.g. when the differential equations are linear combinations of the components of the system). B.Girling in a very interesting lecture described the use of hybrid computers (combinations of analog and digital hardware) for the study of biological systems. The advantages of the analog technique (for example, the rapid solution of differential equations) are married to those of the digital machine (statistical procedures, logical decisions, etc.).

2.4. *Analysis of enzyme kinetic experiments*

The steady-state treatment of simple systems has become highly sophisticated in recent years, due especially to the work of Cleland [7]. This author has derived very useful expressions for the rate laws and has also provided computer programs which extract kinetic constants from a set of data [8]. Programs of the type which iteratively fit the constants to the data until a least-squares minimum is obtained were available in Edinburgh. Perhaps in the future the minimization can be done by hill-climbing procedures, thus avoiding the partial derivatives (see section 3.1) which are sometimes difficult to formulate. One such program for random two-substrate enzymes, has already been written [32]. However, Cleland's approach as well as the related one of Hanson [9] are very convenient in the simpler cases most frequently met with in practice.

2.5. *Derivation of the rate law*

The derivation of the rate law from complicated physicochemical mechanisms, including aggregation of subunits, is usually done with the King-Altman procedure [10] and its modifications [11], sometimes also with the equilibrium method [12–14]. Such methods are very tedious in involved cases, such as, for example, those treated by the Koshland group [15]. Hence, a very promising approach seems to be the computer-programmed derivation of such equations. J.Kutschera described an elegant variant of the published procedures [16,17] which simulates the King-Altman determinant method. The Silvestri and Zahner program [16] was available at the School. It works very rapidly to produce an expression containing up to 10 concentration terms and 30 rate constants.

One should realize that this approach is different from that of the other programs described so far in this essay, since the computer in this case is not doing arithmetic but is instead working out combinations or permutations of algebraic variables. Programs of this type will become important in the future in the field of regulatory kinetics.

2.6. *Physicochemical routines*

A.Grzybowski supplied a set of smaller programs evidently suggested by the practical requirements of a biochemical laboratory. One of them permits the computation of the equilibrium concentrations of Mg-nucleotide complexes, given the initial concentrations and the formation constants of the complexes. A second program calculates pK's from titration measurements in the case of overlapping dissociation regions. Another practical program, submitted by J.H.Leach in a free communication, dealt with ultracentrifuge data processing.

2.7. *Multivariate statistics*

E.F.Harding gave an introduction to this difficult topic, he considered the general type of problems to be solved by advanced statistical methods, which should be included in any high level program. In fact, the estimation of the reliability of the parameters and the validity of the model are usually by-passed by the experimenter, leaving him in doubt when it is necessary to build further experiments on the basis of the model and its parameters. The problem is discussed by Rosenbrock and Storey [18]. J.Barnes discussed a program he has written which provides estimates and confidence limits for the parameters of rate equations, given suitable data.

3. Technical difficulties in numerical mathematics

A naive biochemist with unlimited confidence in

the universal power of mathematics may have been surprised to hear how cumbersome his systems may turn out to be on a computer. A large number of papers and discussions were devoted to numerical problems.

3.1. *Optimization techniques*

Optimization, in biochemical terms, usually means, given a formula for the behaviour of the system and a set of experimental data, finding which parameter values minimize the sum of the squared deviations.

The answer is difficult in the non-linear cases usually met with in practice, especially if restrictions or rigid relationships between the parameters are to be observed. W.H.Swann gave a thorough review of the available techniques [19–23]. Nearly all of them are iterative. One way is to attempt direct minimization simply by trial and error or by more sophisticated procedures. Alternatively, gradient techniques may be applied which calculate partial derivatives (of the function to be optimized or of the least-square function with respect to the parameters to be adjusted) and proceed in the direction so specified. Both concepts have their points (see the review by Swann in this issue), and a choice depends on an "experimental study" (as Michaelson called it) of the numerical behaviour of the actual system.

3.2. *Stiff equations and instabilities*

If differential equations describing biochemical events are to be solved on a computer, approximation procedures are usually applied. R.A.Buckingham gave a lucid description of the common methods [24–27]. The algorithms may be classified as follows: (a) one-step methods, calculating the next solution increment with a formula. The best known are the Euler-Cauchy and the Runge-Kutta methods and their variants; (b) multistep procedures, iteratively improving by a corrector formula the raw increment calculated by a predictor formula, for example the Adams-Basforth method; (c) extrapolation procedures, for instance the Aitken-Neville or the Burlisch-Stoer methods. The idea behind the last named is that some simple formulae give more accurate values if the step length is smaller. Hence an extrapolation to zero step length is possible and yields the best approximation.

In biochemical systems with low accuracy requirements (approx. 0.1%) the single step method seems to be optimal with respect to both computer time and to error. A great danger frequently met with in practice is instability in the numerical procedure; either the approximation becomes worse with increased step length of integration, or the calculated solutions tend to break into spurious oscillations around the correct monotonous value [1]. This greatly decreases the integration speed.

Equations in which these problems are found are described as "stiff"; in general, an equation system is stiff if the time-constants which control the actual solution are much slower than those which return control to the required solution after the variable(s) have been perturbed slightly from their correct values (e.g., by inaccuracy in a numerical integration step). For example, let us consider the following well-known system:

$$E + S \underset{10^6}{\overset{10^6}{\rightleftharpoons}} ES \xrightarrow{10^1} E + P$$

	E	S	ES	E	P
initial concentrations:	10^{-4}	1	0	10^{-4}	0

As is explained in the standard textbooks, this Michaelis-type system rapidly establishes a pseudo-equilibrium with 50% of the enzyme in the free state and the rest as the enzyme-substrate complex. In the pseudo-equilibrium condition 50 concentration units of E combine with S per unit time, but only 0.001% of the ES formed is actually converted to E and P, all the rest is immediately reconverted to E and S! This is a typical stiff system with time constants of 10^{-6} and solution time changes of some 10^{-1}.

It is obvious, without rigorous explanation, that any approximate integration requires millions of calculation steps per minute, since the computer is bound to do the Sisyphean labour of shuttling the E's and S's to and fro. Paradoxically to obtain an accuracy of, say, 1% the computer has to carry more than 10 significant figures through the calculations because, for instance, the following expression is nearly zero:

$$d[ES]/dt \approx 10^6 [E][S] - (10^6 + 10^1)[ES] .$$

An extensive discussion of the stiff equation problem during the meeting resulted in some proposals to improve the situation:

1. D.Garfinkel and B.Hess improve the numerical behaviour of systems they have studied by decreasing the turnover number of enzyme in the stiff parts of the system and proportionally increasing the enzyme concentrations [1]. This trick turned out to have some justification since recent investigations in Hess's laboratory indicate that the actual concentrations of enzyme (as titrated by antibody) are perhaps much greater than is usually assumed.
2. A simple procedure is to take the rate laws instead of the differential equations (i.e., to assume the steady-state). J.Burns's program has implemented this rationale, as already described. A shortcoming is, as just mentioned, that low enzyme concentrations are then assumed which may not be justified.
3. A simple but, in E.M.Chance's opinion, reliable practical check of the solutions is to repeat them with double precision or another step length and to compare the results.
4. A.Curtis recommended a control device for the step length based on the matrix of partial derivatives of the velocities with respect to the concentrations. The disadvantage is that these derivatives have to be calculated.
5. M.Berman proposed the definite use of the law of conservation of mass in biochemical systems to correct deviations from the true solution. D.Garfinkel preferred to use it as an error criterion instead.

4. Compartmentation – an *Asylum Ignorantiae*

V.Moses presented a paper which really is a challenge to the gay optimism which claims that biochemical investigation is advanced enough to allow a consistent mathematical description of complex systems. He produced a vast number of facts, based on tracer experiments, on the assimilation of hexose by different *E. coli* mutants, and his conclusion was simply an appeal to the audience to explain the data without the ominous conjecture of "metabolic compartmentation". In fact this was scarcely possible. No surprise therefore when D.Garfinkel, working on the much more complicated mammalian brain and kidney systems, also found a compartmentation of certain amino acids [28,29]. Finally, B.Hess showed evidence for inconsistencies in the aldolase – triose-phosphate isomerase system of yeast which suggests some type of compartmentation of dihydroxyacetone phosphate [28]. A long discussion followed to find out what compartmentation is actually like. From the sarcastic "compartmentation is when something has gone wrong with the experiment" to the more serious "compartmentation is inconsistency with homogeneous kinetics" or even "a pool is compartmented if the transition probability is homogeneously distributed over its molecule population", all types of personal definitions appeared. This makes the term so confused that in my opinion it would be better avoided altogether.

However, a concept is useful if it suggests experiments. If other than kinetic grounds can be given to establish what the compartmentation in question really means, the vague term may pass as a working hypothesis. Thus Garfinkel supplied arguments that the small amino-acid pool may consist of particles at nerve endings [29]; therefore compartmentation would be what the term already suggests – a spatial diffusion barrier. Hess and Garfinkel assumed a combination of aldolase and triose-phosphate isomerase directly taking the newly formed dihydroxyacetone phosphate from each other without its previous release into the medium. This would be a physico-chemical compartmentation resulting in local concentration peaks of dihydroxyacetone phosphate. If such non-kinetic explanations can really be proved, the concept would be fruitful. On the other hand, a mere cutting up of the system into pieces until they fit the data is only dilettantism.

But why not apply Occam's razor radically? The trouble which crops up with compartmentation is a dialectical consequence of the very introduction of its possibility by assuming homogeneously mixed populations of substances. C.W.Sheppard consequently proposed to avoid the whole formalism of distinct entities with variable properties (such as a pool is) by trying a stochastic approach to the problem [30,31]. A simple example of this is to consider a set of particles moving down a chromatographic separation column. Of course, the book of Nature is written in differential equations with parameters which relate to motion, but what is the use of it here when nobody knows what the parameters mean and what the variables are? In such cases the stochastic approach is a much more realistic way of representing the actual information content of the observables.

5. Envoy

I have tried by simply reflecting the Edinburgh Computer Summer School to give a survey of the scope and the limitations of numerical techniques in Biochemistry. But, on rereading my story, I find that it would be unfair to stress the disadvantages too far and to leave the reader with the discouraging picture of brave chaps constructing huge programs and fighting a hopeless jungle warfare in the labyrinths of numerical methods, trapping biochemists unless they are prepared to crudely distort their systems until they are easily digested by a computer. Such emotive images are not to the point. Of course computer studies take a lot of time, and the outcome is often only a confirmation of what is already suspected. But this is calibration work which is done in any laboratory. The most efficient use of computer studies of large systems is to develop our intuition about the behaviour of systems which are scarcely understood even at a purely qualitative level. Modern biochemistry tends to analyse objects after destruction to the molecular level, in order to explain their behaviour in the integrated state. This abstract resynthesis of large systems evidently requires advanced computing techniques. It is my firm conviction that computation will become an indispensable tool for the biochemist in the near future. FEBS or one of its member societies would be well advised if it were to hold another computer meeting or summer school in about 3 or 4 years. The annual FEBS meetings should from time to time have a special section on computer techniques so as to increase the rate of exchange of information and ideas between European biochemists.

References

[1] D.Garfinkel and B.Hess, J. Biol. Chem. 239 (1964) 971.
[2] E.M.Chance, Computers and Biomedical Research 1 (1967) 251.
[3] D.Garfinkel, J. Biol. Chem. 241 (1966) 286.
[4] A.Kerson, D.Garfinkel and A.S.Mildrran, J. Biol. Chem. 242 (1967) 2124.
[5] M.Berman, Compartmental analysis in kinetics, in: Computers in Biomedical Research, vol. 2, eds. R.Stacy and B.Waxman (Academic Press, New York, 1965) chap. VII.
[6] J.G.Reich, European J. Biochem. 6 (1968) 395.
[7] W.W.Cleland, Biochim. Biophys. Acta 67 (1963) 104.
[8] W.W.Cleland, Advan. Enzymol. 29 (1967) 1.
[9] K.R.Hanson, E.A.Havir and R.Ling, Biophys. Biochem. Res. Commun. 29 (1967) 194.
[10] E.L.King and C.Altman, J. Phys. Chem. 60 (1956) 1375.
[11] S.Cha, J. Biol. Chem. 243 (1968) 820.
[12] J.Monod, J.Wyman and J.-P.Changeux, J. Mol. Biol. 12 (1965) 88.
[13] M.M.Rubin and J.-P.Changeux, J. Mol. Biol. 21 (1966) 265.
[14] D.E.Koshland, G.Nemethy and D.Filmer, Biochemistry 5 (1966) 365.
[15] M.E.Kirtley and D.E.Koshland, J. Biol. Chem. 242 (1967) 4192.
[16] A.J.Silvestri and J.C.Zahner, Chem. Eng. Sci. 22 (1967) 465.
[17] R.O.Hurst, A computer program for writing the rate equations for complex enzymic reactions, Abstracts 5th FEBS meeting, Prague, 1968, Abstract No. 50.
[18] H.H.Rosenbrock and C.Storey, Computational Technique for Chemical Engineers (Pergamon Press, 1966).
[19] M.J.Box, D.Davies and W.H.Swann, Non-Linear Optimization Techniques, I.C.I. Monograph No. 5 (Oliver & Boyd, Edinburgh, 1969).
[20] M.J.D.Powell, Computer J. 7 (1964) 155.
[21] R.Fletcher and M.J.D.Powell, Computer J. 6 (1963) 163.
[22] W.C.Davidon, Variable metric method for minimization, AEC Research and Development Report, ANL-5990 (Rev.).
[23] D.W.Marquardt, J. Soc. Industr. Appl. Math. 11 (1963) 432.
[24] R.A.Buckingham, Numerical Methods (Pitman, London, 1962).
[25] R.Zurmühl, Praktische Mathematik für Ingenieure und Physiker (Springer-Verlag, Berlin, 1965, 4th ed.).
[26] L.Collatz, The Numerical Treatment of Differential Equations (Springer-Verlag, Berlin, 1967).
[27] P.Henrici, Elements of Numerical Analysis (Wiley, London, 1964).
[28] D.Garfinkel, Ann. N.Y. Acad. Sci. 108 (1963) 293.
[29] D.Garfinkel, J. Biol. Chem. 241 (1966) 2918.
[30] C.W.Sheppard, Ann. N.Y. Acad. Sci. 108 (1963) 29.
[31] P.E.Bergner, J. Theor. Biol. 2 (1962) 33.
[32] J.Kowalik and J.F.Morrison, Math. Biosciences 2 (1968) 57.

N.B. Further references are available from the author upon request.