# CONSTRUCTION OF BIOCHEMICAL COMPUTER MODELS

David GARFINKEL
*Philadelphia, Pennsylvania, USA*

This paper described methods for constructing (digital) computer models of biochemical systems when the main object is to investigate the system itself, and not to fit experimental data (e.g., from tracer kinetics) to a set of equations. The author describes model-building as an art which is difficult to communicate, but nevertheless gives valuable tips on the conceptual and practical aspects, from his own considerable experience.

## 1. Reasons for building biochemical models

The biochemistry of systems of physiological significance (multi-enzyme systems, intact cells, etc.) is inherently very complex, as many reactions or processes are occurring simultaneously. The behavior of such systems must often be examined by indirect methods, requiring the manipulation of large amounts of data, which often cannot conveniently be done with pencil and paper. Kinetic data in particular may lead to large numbers of non-linear differential equations. A natural response to this situation is the utilization of computers.

The behavior of a physiologically significant biochemical system is often not readily decomposible into the sum of the known behaviors of its parts, both because of considerable interaction among the parts, and because the parts are commonly studied separately under conditions radically different from those prevailing in the composite system. The properties of even a single multi-enzyme system may look quite different from the sum of the properties of the constituent enzymes; in two recent simulations of glycolysis by the author [1,2] only one of a dozen glycolytic enzymes seemed to be behaving as indicated by the literature. It has recently been shown [3,4] that the conditions under which many enzymes function in the cell invalidate some of the traditional simplifying assumptions of enzyme kinetics, especially the one that a substrate is present in amounts much larger than its enzyme.

Experiments to determine the behavior of complex biochemical systems may be straightforward to describe, but sufficiently lengthy that improved methods of designing experiments would be helpful (e.g., it may require weeks to work up all the enzyme and intermediate determinations from a perfused-organ experiment). Furthermore, the resulting data may be quite difficult to interpret (e.g., radioactive tracer data from a complex metabolic pathway [5]).

Computer assistance with problems of these types often takes the form of an appropriate model. Although physical analogs of biochemical systems are possible, complex systems often lead to mathematical models of sufficient complexity to require computer manipulation. Such models are constructed to quantitatively account for a body of data regarding the system being studied. They can then be used to predict its behavior in a variety of ways (such as seeing how one variable affects another when the actual experiment cannot be performed). The act of construction may also be valuable, as in showing that two pieces of experimental data are actually contradictory when this is not otherwise evident. In particular, constructing a model requires expressing a large volume of data in a compact and unambiguous form.

Biochemical models may be considered part of a sequence extending from genetic models on the one hand to physiological models on the other. One may model the three-dimensional structure of an enzyme [6], the way it is synthesized in the cell [7], its control mechanisms [8,9], and their physiological significance [8,9]; or one may construct a model of a complete cell [10]. Much of the philosophy involved in

building and using these models is quite similar throughout, even though the detailed methodology differs from one area of subject matter to another. In particular, the "biochemical" models described here are quite applicable to membrane transport, as many of the equations for transport are quite similar to their counterparts in enzymology, many of the problems currently being investigated have their biochemical counterparts, and even the economics of the situation may be the same (some transport experiments are sufficiently long and tedious that assistance in designing them would be worthwhile).

It may reasonably be claimed that biochemical models have reached a stage of maturity where they are beginning to provide useful biochemical information, and are likely to become the best way of obtaining certain types of biochemical information. This will be particularly important as biochemistry, having decomposed important systems into constituent parts and studied their properties, now starts to see how the parts fit together. Biochemical models should be even more useful in manipulating complex biochemical systems, as when one tries to get a pathway to make a product it was not intended to make (e.g., for industrial purposes). Direct experimentation here may lead only to unexplainable failure; failure of a model can be accompanied by a detailed written report of exactly how and why it failed.

## 2. Conceptual aspects of building biochemical models

The process of building a model consists of assembling a collection of ideas and approximations representing the appropriate subject matter, and when this proves to be inadequate (as it almost always does at the first try), adding additional ideas and finding better approximations. Models tend to grow by accretion, and to become more refined and sophisticated with the passage of time, but will not necessarily come to fit the data more accurately. In the author's experience this process has been limited by either the generation of new ideas or the acquisition of additional experimental data.

A model should be kept as simple as possible; the more complex it becomes, the fuzzier the conclusions obtained from it will be. However, the other side of this coin is that it must not be so simple as to oversimplify the system represented. There is no need to have all parts of a model of equal simplicity; in particular one might very well wish to examine something in great detail, and have the other parts of the same system (acting as an "environment") much less detailed.

If the subject area being modelled is clearly divided into sub-systems (e.g., a multi-enzyme system) these may be separated and modelled separately, which is probably easier than modelling the composite system in its entirety. It is possible to make fairly standard sub-system models that need be done only once and are usable repeatedly, even being exchangeable among different users, although it should be cautioned that these will not necessarily meet everyone's needs.

The uniqueness of models is a continuous source of concern. Often two or more alternative models may be constructed for a given system, each representing the statistically best-fitting version (among an infinite number of variants) of a given collection of ideas and approximations. Some method of choosing among them (or combining them) is necessary. Choosing the simplest ("Occam's razor") is a possibility, as is choosing the one that explains the most facts. Choosing the best-fitting to the data probably is not desirable if the fit is only slightly better (it should be significantly better). Preferably the choice should be made on other than statistical grounds, especially by designing a critical experiment, as by finding the conditions where the differences between the consequences or predictions will be the greatest.

## 3. Limitations of biochemical models

It should be kept firmly in mind that a successful model, which is equivalent to a quantitative theory, does not constitute a final proof of anything. Final proof, as far as it is possible, must come from experiment (often experimental and theoretical developments may alternate for a long time before a given problem is finally settled). However, an unsuccessful attempt to build or fit a model can be a disproof; if a given theory cannot be successfully modelled, then it is unlikely to be correct. Often such disproof can be very rapid; it has been the author's experience that if one cannot obtain approximately the right behavior of a model after adding a new idea in half a dozen or

so tries, where the parameters involved may be varied over rather wide ranges, then it is unlikely that success will be obtained by the insertion of other numerical values. (This generalization is based only on accumulated experience; there is no semblance of a formal proof for it and doubtless many exceptions exist). Since models inherently suggest, confirm, or disprove possibilities rather than offering conclusive proof, there is often difficulty in publication: they are unpopular with biochemical editors who prefer to print communications that appear to conclusively prove things.

It should be stressed that model building cannot be done in isolation from experiment; nor will it tolerate a lack either of familiarity with the subject matter being modelled, or of the appropriate experimental data. The quality of the model will be no better than the quality of the data on which it is based. This is not quite so stringent when a relatively theoretical question is investigated, when minimal interaction with experiment may be sufficient to keep the investigation going in the right direction. It is possible for a person with experience primarily in model-building to subsequently learn the appropriate subject matter, as well as the reverse.

A model is built *with* a computer, but not *by* a computer. The computer will tell the user what the properties of a given model are and it can be made to systematically manipulate the model, as by adjusting parameters. At the present state of technology the computer will not automatically build a model or anything of the sort. This may perhaps change somewhat in the future, especially when an appreciable library of models has been constructed and one may search it for analogies. A partial step in this direction is to automatically test assumptions in some situations involving missing data where the desired behavior of the model is known by having the computer calculate the simulation input on the basis of these assumptions, and then doing the simulation and seeing how it works.

## 4. Methodology and technology of constructing biochemical computer models

This is actually something of an art and in part must be learned by experience. The author has not yet discovered any straightforward method of describing how it is done, and has found communication regarding it rather difficult, even with a person of suitable background with whom communication is otherwise easy. The formal discipline that perhaps best serves as preparation for it is chemical engineering, as constructing a model is not very different from designing a chemical process or its equipment.

Biochemical models may be constructed with either analog or digital computers. The analog has the advantages of simplicity, lower initial cost, and ease of operation, and is often faster for this type of work. Indeed, the first biochemical simulation was performed with an analog computer 25 years ago [11], long before digital computers existed. On the other hand most analogs are strictly limited in size (leading to serious over-simplification of many biochemical problems) and also less amenable to comparison of experimental and calculated results. The capacity of the larger digital computers now in common use is easily sufficient for the most complex models now being studied. The author generally recommends that an analog computer be used when this is both possible and convenient. A hybrid computer (analog + digital) might be even better when available; this combines the flexibility of the digital with the speed – and size limitations – of the analog.

If a digital computer is used, it may either be a large central batch-processing computer *, or one having reactive interactions with the user, as by time-sharing *. The latter arrangement is definitely preferable; some people [12] feel that it is almost mandatory and that one cannot do good simulation work with a batch-processing machine, although this is probably too pessimistic. In particular, since constructing a model is very often a trial-and-error process, it is helpful to be able to see the progress of each trial, so as to be able to change it or stop it if it is unsatisfactory.

Programs are available both to construct biochemical models on large digital computers and to perform

* A batch-processing computer processes many jobs together in a batch; these are received from and returned to the user by the computer operators, so that the user has no direct contact with the computer. Most university computers are presently of this type. A time-sharing machine is able to interact with a user (usually at his own remote console) in a conversational fashion, by rapidly alternating its attention among many users. Some computers presently function in both modes.

some of the incidental calculations needed in the process. There exist many general-purpose simulation languages [13,14] which are applicable to this type of problem, but they usually have the disadvantages of requiring the user to write his own differential equations and of sacrificing efficiency for the sake of generality. In addition there are specialized languages (often available for distribution to interested users) intended primarily for constructing biochemical models [15–18]. The author's contribution to this group [18] is a machine-independent program intended to permit the user to describe his model in biochemical language (and to exchange such models with other users) while the computer handles the mathematics.

Attempts are being made [19,20] to establish a standard form of what a model should be like, how it should be described and documented, as well as to find out how efficient different computers are in computing its behavior. This is intended in part to facilitate communication of information about models, which may be difficult owing to their complexity. In building a model one has to keep track of why one made the decisions one did, what mistakes were made, what was learned along the way, etc. It may be desirable in describing a model to also include its unsuccessful predecessors, and the reasons for their failure to fit the data or other inadequacies.

The mathematics involved in computing the behavior of a biochemical model is not particularly exotic, often consisting primarily of the solution of non-linear ordinary differential equations as a function of time. However, these equations display the troublesome property called "stiffness" (which slows computation), usually as a result of the very large numerical values for naturally occurring rate constants that must be inserted into them. There is no easy solution for this difficulty, but it may sometimes be alleviated by representing the velocity of an enzyme by the appropriate algebraic rate law and thus eliminating from the system the differential equations which are causing the most difficulty.

It should be noted that simulating biochemical systems is rather different from simulating situations, the subject area where the art is most advanced. Business situations usually involve discrete rather than continuous events, input numbers are more accurately determined, often with no experimental error, and the object of the simulation is usually some kind of optimization, either by maximizing efficiency or by increasing profit, rather than to understand how the system being simulated is in fact organized and controlled.

Some of the non-computing aspects of constructing a model also deserve comment. For instance, one often has relatively few experimental points, but a large amount of background material (such as the activity of a given enzyme in a given organ under certain conditions), usually obtainable from the literature, which relates these experimental points and provides a setting for them. This situation has not been well analyzed statistically and there is no standard procedure for it. Here in particular expert knowledge of the subject matter is required.

## 5. Probable future developments of biochemical models

The use of this technique now appears to be entering a period of exponential growth, as more and more people become interested in it. In particular economics presently seems to favor its further growth, as the cost of doing a given computation is rapidly decreasing, whereas the cost of doing a given experiment is perhaps slowly increasing. It is even possible that in the foreseeable future the cost of the computer time needed to build a given biochemical model will decrease by several orders of magnitude (although this will doubtless be partially compensated for by increasing complexity of models).

It seems probable to the author that some of the rapidly increasing amount of quantitative biochemical information available will probably come to be grouped and organized in standard models for important tissues, such as mammalian liver and heart, yeast, *E. coli* $K_{12}$; once constructed these will continue to be added to and refined for an appreciable period of time, and hopefully will be widely used.

It is hoped that this technique will become a standard technique of biochemistry, of special value for handling complex situations.

Written during the tenure of a Research Career Development Award (E3-GM-5469) and primarily supported by grant FR-15, both from the National Institute of Health.

## References

[1] D.Garfinkel, R.Frenkel and L.Garfinkel, Computers and Biomedical Research 2 (1968) 68.
[2] M.J.Achs and D.Garfinkel, Computers and Biomedical Research 2 (1968) 92.
[3] B.Hess, in: Advances in Enzyme Regulation, ed. G.Weber (Pergamon Press), in the press.
[4] P.A.Srere, in: Metabolic roles of citrate, Biochemical Society Symposium No. 27, ed. T.W.Goodwin (Academic Press, London) p. 11.
[5] M.Haut and S.Basickes, Computers and Biomedical Research 1 (1967) 139.
[6] C.Levinthal, Sci. Am. 214, No. 6 (1966) 42.
[7] F.Heinmets, Analysis of Normal and Abnormal Cell Growth (Plenum Press, New York, 1966).
[8] D.Garfinkel, J. Biol. Chem. 241 (1966) 286.
[9] O.A.Kerson, D.Garfinkel and A.S.Mildvan, J. Biol. Chem. 242 (1967) 2124.
[10] R.Weinberg, Bact. Proceedings (1968) p. 38; Abstracts, 6th Annual Symposium on Biomathematics and Computer Science in the Life Sciences, Houston, 1968, p. 102.
[11] B.Chance, J. Biol. Chem. 151 (1943) 553.
[12] National Institutes of Health Physiology Training Committee, in: The Physiologist 11 (1968) 115.
[13] D.Teichrow and J.F.Lubin, Comm. Assoc. Computing Machinery 9 (1966) 723.
[14] D.Teichrow, J.F Lubin and T.D.Truitt, Simulation 9 (1967) 181.
[15] D.F.DeTar and C.E.DeTar, J. Phys. Chem. 70 (1966) 3842.
[16] M.Berman, E.Shahn and M.F.Weiss, Biophys. J. 2 (1962) 275.
[17] E.C.DeLand, Rand. Corp. Memorandum RM-5405-PR (1967).
[18] D.Garfinkel, Computers and Biomedical Research 2 (1968) 31.
[19] J.McLeod, Simulation 7 (1966) 324.
[20] J.McLeod, Simulation 10 (1968) 37.