

Computer Simulation of Biochemical Kinetics

Pedro Mendes & Douglas B. Kell

**Institute of Biological Sciences, Edward Llwyd Building, University of Wales, Aberystwyth,
SY23 3DA, United Kingdom**

<http://gepasi.dbs.aber.ac.uk/home.htm>, prm@aber.ac.uk, dbk@aber.ac.uk

Fax: + 44 1970 622350, Phone: + 44 1970 622353

INTRODUCTION

The description of the kinetics of biochemical (and of course chemical) reactions can be made on two levels. At a basic, molecular, level a reacting chemical system can be described as an ensemble of particles that react on collision. Each pair of particles is assigned a probability for collision per unit time and volume. The evolution of the number of particles of each species is represented by a stochastic time series. At a macroscopic level, the various chemical species are represented by their concentrations in the reactor. The evolution of the various concentrations with the progress of reaction is described in terms of differential equations. These are partial differential equations if the spatial dimension (diffusion and convection processes) is taken into account, or ordinary ones if time is the only independent variable considered. Both these theoretical frameworks for the kinetics of chemical reactions were known before the appearance of computers. However their application to biochemical systems was quite limited.

To simulate a biochemical reaction at the (stochastic) molecular level one has to carry out a very large number of operations. In principle the macroscopic (deterministic) representation can be dealt with using infinitesimal calculus; however due to the non-linearity of the differential equations for biochemical systems, analytic solutions are not possible. Alternatively, there are methods for calculating numerical solutions but these, as with the stochastic representation, require large numbers of operations to be carried out. The impossible problem of executing large numbers of such calculations is bound up both with the inevitability of human errors and the length of time required (in many cases estimated to be larger than the average human lifetime). The deterministic representation was nevertheless used since the early days of biochemistry, as applied to the study of (single) enzyme kinetics. This was possible because of the use of approximations (for example the equilibrium assumption of Michaelis and Menten [1]).

In this paper, we review the use of computers in the simulation of systems of biochemical reactions. Due to space limitations we shall deal only with the deterministic approach with a stronger emphasis on the computational simulation tools. We have opted for citing only a few works where a more complete list would increase the size of this paper beyond the acceptable limit. In doing so, we have tried to select works which are as diverse as possible.

ANALOGUE COMPUTERS

One of the earliest simulation studies of biochemical reactions was carried out by Chance still during the second world-war. In a remarkable paper [2] where experiment, analysis and simulation were all used, Chance showed the validity of the Michaelis-Menten model for the action of the enzyme peroxidase. The simulations were carried out with a mechanical differential analyser. The

differential analyser belongs to a class of computers known as analogue. These computers were designed to take input in terms of continuous variables, such as hydraulic pressure or voltage. These were then processed by various units, which altered them according to particular mathematical functions (including derivatives and integrals). Analogue computers were programmed by coupling various elementary devices, and in our field of interest they would solve a set of differential equations, in a continuous simulation of the model. Note that unlike digital computers, there is no round-off error involved in calculations processed with analogue computers.

The mechanical differential analyser was made obsolete by electronic analogue computers. These were used mostly in the 1950s and 60s (e.g. refs 3-7) but also later [8]. Electronic analogue computers were faster than their mechanical equivalents, they could be coupled with cathode ray tubes for real-time visualisation and were also easier to set up (even though they still required knowledge of electronics). On the other hand, analogue computers have limits on the size of the models that they can handle (these are practical limits connected with the space taken by the computer and the number and complexity of the connections between the various modules).

MAINFRAME DIGITAL COMPUTERS

Digital computers were also developed during the second world war and after its end they were very quickly adopted for non-military use. The main advantage of digital computers over their analogue counterparts is that one machine can perform many unrelated tasks without any need to change the connections of the basic hardware components. Because these computers (mainframes) were quite expensive, they had to be shared between many research groups. Processing was done in batches which meant that writing and testing programs took a very long time. However, this class of computers has been used for biochemical simulations since the early days. In terms of their application to biochemical models, digital computers had the additional advantage over their analogue counterparts that they could handle much larger models (more enzymes and metabolites) [9].

There is a large body of publications describing biochemical simulations executed by digital mainframe and mini computers (for example [9-15]). Some of these simulations were executed with programs written specifically for the models under study. However some other programs were written that were not specific for a single model. These generic simulators have the obvious advantage that once written, they can be used for more than one model. The group of David Garfinkel was a pioneer with their BIOSSIM simulator [16, 17]. With BIOSSIM each model was defined using a series of statements, this was then processed by a program that translated the input to the differential equations (written in FORTRAN) and by another that would solve them to determine the time course of the various biochemical concentrations and fluxes. Other programs similar to BIOSSIM were described in the literature around this time [18-20].

During this period of large mainframe computers, one technical problem had to be overcome in order to make generic simulators cope with all biochemical reaction models. It became obvious that these models are more often than not composed of systems of stiff ordinary differential equations (ODE) [21,22]. Stiffness is a property that arises by the co-existence of kinetic constants of very different orders of magnitude. It is a problem because it requires integration step sizes so small that normal numerical methods would (i) require massively large amounts of execution time, and (ii) accumulate such a level of truncation error that their results would be meaningless [23]. Fortunately there are methods well suited for handling systems with this characteristic and these were therefore incorporated in some of the generic simulation programs [17, 20].

All of the general-purpose metabolic simulators created in this era were “language-driven”, i.e. the user described the model with a series of statements that used biochemical terms, but that were

otherwise very similar to computer languages like FORTRAN and ALGOL 60. The operation of these simulators involved a step of translation of the “biochemical program” to a FORTRAN program (or other computer language), that was consequently compiled and only then ran to produce results. This need for compilation did not constitute a problem since mainframe computers were invariably supplied with high-level language compilers. Some of these programs [16,17] were portable, in the sense that they could be made to run on various computers (different underlying hardware). However this involved quite some effort; for example it took a minimum of one week for BIOSSIM to be made to run on a computer different from the one for which it was created [17].

PERSONAL COMPUTERS AND WORKSTATIONS

By the end of the 1970s and as a consequence of the level of miniaturisation of electronic circuits achieved then, small computers started to become available at a price that was affordable to individuals. These microcomputers were at first only really useful to the electronics hobbyist, but soon they started becoming used in all areas of human activity. Science was no exception, and at a time when research budgets were starting to suffer cuts cheap computers could only be a good thing. Another class of computers that became popular at this time are known as workstations. These were the direct evolution of mini-computers like DEC’s PDP series. Workstations were cheap enough to be bought by departments rather than universities yet, they were much more powerful than the personal computers then available.

This is the age of BASIC, a computer language that does not need to be compiled. The use of BASIC made it evident that the way in which programs interacted with the user was very cumbersome. Progress can be made much faster if one has an immediate reply from the machine, instead of passing through several compilation stages and batch runs. The motto of this age (which still continues to this date) was user-friendliness.

In terms of the simulation of biochemical systems a curious phenomenon happened. While the number of people using computers to simulate biochemical processes increased markedly, the knowledge acquired previously by groups such as that of Garfinkel, seems to have vanished into a void. The only exception is the program SCoP [24] for micro and minicomputers, which is very similar to BIOSSIM in its operation and capabilities (including having to pass the “biochemical program” through two compilers but also incorporating methods for stiff ODEs). Except for this honourable case, most of the problems of metabolic simulation and the techniques to solve them were slowly rediscovered through the 1980s by those writing metabolic simulators for microcomputers.

Microcomputers had many advantages, but there was one that is only being partially overcome at present. Because there were many manufacturers, operating systems and compilers these computers were mostly incompatible with each other. This also slowed progress as whenever a new machine appeared new programs had to be written specifically for that. More often than not these programs would not even be written (to my knowledge there is no generic metabolic simulator for the Commodore Amiga, for example).

Hofmeyr was a pioneer in this area of microcomputer software for metabolic simulations. His program METAMOD [25] was written for the BBC Model B. This program only dealt with steady-states but could apply metabolic control analysis (MCA) [26,27], a theoretical framework for the analysis of control (for a recent review see ref. 28). METAMOD was an interactive menu-driven program. Another program that would later on be ported to microcomputers [29, 30], SCAMP [31] was originally written for a minicomputer (Prime) in 1986. SCAMP is capable of dealing with both time-course simulations and steady-state analysis (including MCA). However this program follows the old style of taking input from a program written in a specific language that has to be compiled

(the compilation step is optional in the most recent version of this program for IBM-PCs, Atari, and Unix computers).

As the microcomputer industry started maturing, two types of microcomputers became standards: those compatible with the IBM-PC and the Apple Macintosh. While there is only one general-purpose metabolic simulator for the Macintosh, there are many for the IBM-PC. ESSYNS [32] is an MS-DOS program that simulates both reaction progress and steady states, using power-law approximations of the kinetic equations of each metabolic step [33]. This simulator is capable of determining the stability of the steady-state solutions and apply the BST analysis of Savageau [33], in many respects similar to MCA. MetaModel [34] is similar to METAMOD in its capabilities (and to some extent based on it) but runs under MS-DOS. Gepasi, originally an MS-DOS program [35], was later converted to Microsoft Windows [36] and is now a 32-bit MS Windows program (See Mendes, elsewhere in this volume). Gepasi, like SCAMP, calculates steady states as well as reaction progress and characterises the steady states with MCA. It is, however, an interactive program using the array of “controls” of MS Windows (buttons, etc.) and contains a context-sensitive manual (“help file”). SIMFIT [37] is an MS-DOS program that simulates reaction progress and can also fit model parameters to experimental data using a least-squares algorithm. METASIM [38] was an attempt to create a representation of metabolic systems using an object-oriented paradigm and computer language. This representation had the drawback that it limited the ODE integration algorithm to be the Euler method, which is unfortunately too primitive to be able to handle any realistic metabolic model. This line of research seems to have been abandoned, even though there are some simulators that were written in object oriented languages (Gepasi version 3 is written in C++). KINSIM [39] is available for VAX minicomputers, IBM-PC and Apple Macintosh, it is a menu-driven program, capable of simulating reaction progress and to perform least-squares fits of parameters to experimental data. MIST [40] is a MS Windows based simulator that can also calculate reaction progress, steady states and use MCA for characterisation of the steady states. Those programs available via the Internet may be accessed from our home page (<http://gepasi.dbs.aber.ac.uk/home.htm>).

Because microcomputers are used in almost all human activities now, there is an enormous amount of software available. A certain class of general-purpose simulation programs, which were primarily designed for other purposes, can be applied to create metabolic models. These programs lack a vocabulary that is oriented towards biochemical problems and require the user to have a deep theoretical knowledge about simulation of biochemical reactions. However, there is a realistic danger that this type of programs could be misused (for instance it is unfortunate to see concepts like “variables” and “parameters” being confused in the literature so often [41]). Without trying to be exhaustive, some of these programs are: Mathematica by Wolfram Research (for many different hardware platforms) which is a general mathematics package, including symbolic and numerical manipulations; Matlab by The Mathworks, Inc. (also for many hardware platforms) very similar to Mathematica; MLAB by Civilized Software, Inc. (for IBM-PC and some Unix workstations), also similar to Mathematica; Stella, by High Performance Systems (for the Apple Macintosh) which is an object-oriented generic modelling system, with a very appealing front-end; and SPICE (for many hardware platforms) which is really an electrical circuit simulator. The latter can be used to build a simulation of an analogue computer that itself simulates a biochemical system. Not only does the prospective user need to be an expert on metabolic simulation, but also on electronics (as in the 1950s!).

FUTURE PROSPECTS

Kootsey [42] discussed the possibility of a client-server approach to the simulation of biochemical systems. In this approach the simulation engine runs on a remote (high specification) computer, while the front-end runs on a client (less-powerful) computer. This architecture allows for various

clients to access the server simultaneously and the server could eventually process several simulations concurrently (parallel or distributed processing). With the expansion of the Internet this approach seems even more plausible today. In fact the protocol for interfacing the client with the server could easily be based on the world-wide-web (see Mendes & Kell, elsewhere in this volume), using HTML forms or the emerging JAVA script language.

Even though optimisation has been a feature included in some generic metabolic simulators [20, 24, 37, 39, 43], its application has been rather modest and has been confined mostly to parameter fitting and system identification. Optimisation methods can be useful for other purposes. For example optimisation can be applied to models of known structure and parameter values for which one wants to find sets of parameter values that would produce optimal values of some variable (for example maximisation of a metabolic flux of interest). This application of optimisation methods has indeed only been described in a few papers [44,45] and was never done through a general-purpose simulator. Indeed, there is particular interest in the inverse problem of metabolism, in which one wishes to do this for models in which one knows only the structure [46]. Another application of optimisation principles is in the simulation of evolutionary processes [47]. The use of optimisation methods in metabolic modelling for the purposes of metabolic engineering looks as if it may well become a major topic. Another class of methods that can be applied in a similar manner to optimisation are continuation methods [48]. These are used to obtain steady-state solutions of a model based on a previously calculated steady state for a set of parameters that is close to the required one. This technique is especially useful for models whose solutions bifurcate with changes in one (or more) of its parameters. Only a couple of general-purpose metabolic simulators have incorporated continuation methods [24, 30]. The software package AUTO by Doedel [49] is a set of FORTRAN routines that allows one to carry out continuation (bifurcation) analysis on a set of ODEs. More recently Khibnik et al. [50] developed LOCBIF, an interactive program for essentially the same purposes as AUTO. Both these packages require the user to input ODEs rather than the biochemical reactions. Continuation methods can be very useful for identification of all possible steady-states for a specific set of model parameters.

Apart from some scattered works (for example references 51-54), metabolic simulation studies have considered only one compartment and assumed that it was well-stirred so that diffusion effects were negligible. The simulation of models with more than one compartment is not hard to implement in a generic simulation program (see Mendes, elsewhere in this volume), but the inclusion of diffusion effects is more problematic. The computational methods for reaction-diffusion models are not complex, from an algorithmic point of view, but require large amounts of memory and fast processors. Until recently such simulations could only be executed in super-computers (except for very small models of one or two spatial dimensions) but the increasingly cheaper prices and more powerful specifications of microcomputer hardware mean that we are reaching a stage in which it is feasible to carry out such computations on modern desktop computers. It would be of great interest to see more studies of realistic reaction-diffusion metabolic models, since this would greatly increase our understanding of cellular processes [55].

If the past of this field is to be trusted as an indicator of what the future will be, one must be cautious not to be over-enthusiastic. Many advances made in the somewhat distant past are still not being used routinely today (e.g. pattern recognition [56]) so the issues discussed in the above paragraphs might not materialise. In contrast to the use of computers that have been booming in other areas of biological research (such as DNA and protein sequence processing) the simulation of biochemical reaction systems has still to catch up with experimentalists. This can only happen if generic metabolic simulators continue to progress in their capabilities and, most importantly, in their ease of use and wide dissemination. Unfortunately, even though there are a number of generic and user-friendly metabolic simulators, as described in the previous section, some authors seem to be unaware of their existence [57]. It is of no use to adopt the view that experimentalists will never

do simulation because it involves the understanding of some hard bits of algebra and calculus. Metabolic simulation can only become a common activity of those doing experimental research if we make an effort to make good simulators that are easily available, and carry out and publish studies that integrate experiment and simulation side by side (e.g. ref. 58).

Microcomputers are at present barely distinguishable from workstations in terms of processing speed and memory capacity, and prices of both continue to fall. There is now the possibility of incorporating more features in future metabolic simulators, these should enhance the way in which biochemists who are not computer specialists interact with such programs. We may also incorporate methods that would be unthinkable five years ago (at least for this class of computers). Communications between researchers have been greatly enhanced with the advent of the Internet, which is already used to distribute software, data and ideas. The benefits of the Internet may be even greater in the near future, maybe allowing simulators to be used remotely. We live in exciting times.

ACKNOWLEDGEMENT

We are indebted to the Chemicals and Pharmaceuticals Directorate of the UK BBSRC for financial support.

REFERENCES

- [1] Michaelis, L. & Menten, M.L. (1913) *Biochem. Z.* 49, 333-369.
- [2] Chance, B. (1943) *J. Biol. Chem.* 151, 553-577.
- [3] Chance, B., Greenstein, D.S., Higgins, J. & Yang, C.C. (1952) *Arch. Biochem. Biophys.* 37, 322-339.
- [4] Chance, B., Williams, G.R., Holmes, W.F. & Higgins, J. (1955) *J. Biol. Chem.* 217, 439-451.
- [5] Hempling, H.G. (1967) *Biochim. Biophys. Acta* 135, 355-358.
- [6] Noat, G. & Ricard, J. (1968) *Eur. J. Biochem.* 5, 71-72.
- [7] Higgins, J. (1964) *Proc. Natl. Acad. Sci. USA* 51, 989-994.
- [8] Lassmann, A. & Rietbrock, N. (1982) *Naunyn-Schmiedeberg's Arch. Pharmacol.* 320, 189-195.
- [9] Chance, B., Garfinkel, D. Higgins, J. & Hess, B. (1960) *J. Biol.Chem.* 235, 2426-2439.
- [10] Garfinkel, D. & Hess, B. (1964) *J. Biol. Chem.* 239, 971-983.
- [11] LeFevre, P.G. (1966) *Biochim. Biophys. Acta* 120, 395-405.
- [12] Kerson, L.A., Garfinkel, D. & Mildvan, A.S. (1967) *J. Biol. Chem.* 242, 2124-2133.
- [13] Goldbeter, A. & Lefever, R. (1972) *Biophys. J.* 12, 1302-1315.
- [14] Turner, Illingworth, J.A. & Tipton, K.F. (1974) *Biochem. J.* 144, 353-360.
- [15] Rapoport, T.A., Heinrich, R. & Rapoport, S.M. (1976) *Biochem. J.* 154, 449-469.
- [16] Garfinkel, D. (1968) *Computers Biomed. Res.* 2, 31-44.
- [17] Roman, G. & Garfinkel, D. (1978) *Computers Biomed. Res.* 11, 3-15.
- [18] Illingworth, J.A. (1972) *Biochem. J.* 130, 45p.
- [19] Park, D.J.M. & Wright, B.E. (1973) *Comput. Progr. Biomed.* 3, 10-26.
- [20] Curtis, A.R. (1976) *Biochem. Soc. Trans.*, 4, 364-371.
- [21] Hemker, P.W. (1972) in *Analysis and simulation of biochemical systems.* (Hemker, H.C. & Hess, B eds.) Amsterdam, North-Holland, pp. 59-80.
- [22] Garfinkel, D., Marbach, C.B. & Shapiro, N.Z. (1977) *Annu. Rev. Biophys. Bioeng.* 6, 525-543.
- [23] Gear, C.W. (1971) *Numerical initial value problems in ordinary differential equations.* Englewood Cliffs, Prentice-Hall.
- [24] Kootsey, J.M., Kohn, M.C., Feezor, M.D., Mitchell, G.R. & Fletcher, P.R. (1986) *Bull. Math. Biol.* 48, 427-441.
- [25] Hofmeyr, J.H.S. & van der Merwe, K.J. (1986) *Comput. Applic. Biosci.* 4, 243-249.

Published as: Mendes, P. & Kell, D. B. (1996). Computer simulation of biochemical kinetics. in *BioThermoKinetics of the living cell* (ed. H. V. Westerhoff, J. L. Snoep, F. E. Sluse, J. E. Wijker & B. N. Kholodenko), pp. 254-257. BioThermoKinetics Press, Amsterdam.

- [26] Kacser, H. & Burns, J.A. (1973) *Symp. Soc. Exp. Biol.*, 27, 65-104.
- [27] Heinrich, R. & Rapoport, T.A. (1974) *Eur. J. Biochem.*, 42, 89-95.
- [28] Fell, D.A. (1992) *Biochem. J.* 286, 313-330.
- [29] Sauro, H.M. & Fell, D.A. (1991) *Math. Comput. Model.* 15, 15-28.
- [30] Sauro, H.M. (1993) *Comput. Applic. Biosci.* 9, 441-450.
- [31] Sauro, H.M. (1986) PhD thesis, Oxford Polytechnic.
- [32] Irvine, D.H., Voit, E.O. & Savageau, M.A. (1991) in *Canonical Nonlinear Modeling. S-system approach to understanding complexity.* (Voit, E.O. ed), New York, Van Nostrand Reinhold, pp. 133-141.
- [33] Savageau, M.A. (1976) *Biochemical systems analysis. A study of function and design in molecular biology.* Reading, Massachusetts, Addison-Wesley.
- [34] Cornish-Bowden, A. & Hofmeyr, J.-H.S. (1991) *Comput. Applic. Biosci.*, 7, 89-93.
- [35] Mendes, P. (1994) in *Biothermokinetics.* (Westerhoff, H.V. ed), Intercept, Andover, pp. 231-233.
- [36] Mendes, P. (1993) *Comput. Appl. Biosci.* 9, 563-571.
- [37] Holzhütter, H.G. & Colosimo, A. (1990) *Comput. Appl. Biosci.* 6, 23-28.
- [38] Stoffers, H.J, Sonnhammer, E.L.L, Blommesteijn, G.J.F., Raat, N.J.H. & Westerhoff, H.V. (1992) *Comput. Appl. Biosci.* 8, 443-449.
- [39] Frieden, C. (1993) *Trends Biochem. Sci.* 18, 58-60.
- [40] Ehrlde, M. & Zacchi, G. (1995) *Comput. Applic. Biosci.* 11, 201-207.
- [41] Kell, D.B. & Sonnleitner, B. (1995) *Trends Biotechnol.* 13, 481-492.
- [42] Kootsey, J.M. (1986) *Bull. Math. Biol.*, 48, 405-415.
- [43] Curtis, A.R. & Chance, E.M. (1972) in *Analysis and simulation of biochemical systems.* (Hemker, H.C. & Hess, B eds.) Amsterdam, North-Holland, pp. 39-57.
- [44] Gilman, A. & Ross, J. (1995) *Biophys. J.* 69, 1321-1333.
- [45] Torres, N.V., Voit, E.O. & González-Alcón, C. (1996) *Biotech. Bioeng.* 49, 247-258.
- [46] Mendes, P. & Kell, D.B. (1996) *BioSystems* 38, 15-28.
- [47] Bray, D. & Lay, S. (1994) *Biophys. J.* 66, 972-977.
- [48] Ficken, F.A. (1951) *Comm. Pure Appl. Math.* 4, 435-456.
- [49] Doedel, E.J. (1981) in *Proceedings of the 10th Manitoba Conference on Numerical Mathematics and Computers*, Winnipeg, Manitoba, Canada, Univ. Manitoba Press, pp. 265-284.
- [50] Khibnik, A.I., Kuznetsov, Y.A., Levitin, V.V. & Nikolaev, E.V. (1993) *Physica D* 62, 360-371.
- [51] Goldbeter, A. (1973) *Proc. Nat. Acad. Sci. USA* 70, 3255-3259.
- [52] Fell, D.A. (1980) *J. Theor. Biol.* 84, 361-385.
- [53] Goodwin, B.C. & Trainor, L.E.H. (1985) *J. Theor. Biol.* 117, 79-106.
- [54] Marmillot, P., Hervagault, J.-F. & Welch, G.R. (1992) *Proc. Nat. Acad. Sci. USA* 89, 12103-12107.
- [55] Mendes, P., Kell, D.B. & Welch, G.R. (1995) *Adv. Mol. Cell Biol.* 11, 1-19.
- [56] Garfinkel, D., Polk, W., Higgins, J.J. & Ochsner, R.T. (1962) *Comm. A.C.M.*, 5, 115-118.
- [57] Jackson, R.C. (1995) *Toxicol.* 197-205.
- [58] Schuster, R. & Holzhütter, H.G. (1995) *Eur. J. Biochem.* 229, 403-418.