# Metabolic modeling: a tool of drug discovery in the post-genomic era

Eberhard O. Voit

**The Human Genome Project, DNA microarrays, proteome chips and metabolic profiles, among others, are generating unprecedented amounts of valuable information about the genetic and metabolic responses of organisms to stimuli. This wealth of information poses a great scientific challenge, namely the development of novel, effective methods for functional analysis and interpretation. It is proposed here that biomathematical systems models must accompany data generation and management. A particularly effective framework for this purpose is canonical modeling based on Biochemical Systems Theory. The key concept of this theory is the formulation of biological phenomena as systems of differential equations, in which all processes are represented as products of power-law functions.**

**Eberhard O. Voit**
Department of Biometry and Epidemiology
and Department of Biochemistry and Molecular Biology
Medical University of South Carolina
303 K, Cannon Place
135 Cannon Street
Charleston, SC 29425, USA
tel: +1 843 876 1122
fax: +1 843 876 1126
e-mail: VoitEO@MUSC.edu

▼ Drug discovery through trial and error has been successful throughout the centuries. Treatments with positive effects were retained and unsuccessful remedies were discarded. Both Western and Eastern medicine attest to the power of this strategy. Nonetheless, with the advent of organic chemistry and pharmacology came new requirements. Pharmaceutical companies had to provide explanations for the positive effects of new drugs and these explanations required scientifically rigorous testing of biological mechanisms. Physiological causes for specific diseases were discovered and, in some cases, causes for these causes. Encouraged by these successes, a scientific deduction was made beyond drug development, namely that, to understand physiological functioning or disease, an organism had to be disassembled and its components studied, as well as the components of components, until the ultimate building blocks of life were reached. These would reveal the fundamental mechanisms of health and disease.

## Reduction and reconstruction

The scientific paradigm of focusing on smaller components at deeper levels of biological organization is now called reductionism, and has been extremely successful over the past decades and will, without doubt, continue to be successful. Completion of the Human Genome Project has, in some sense, meant reaching the goal of reductionism because information about health, disease and the body's response to stimuli and drugs has been reduced to a string composed of just four different nucleotides.

Although an incredible feat of the human mind, the unprecedented accumulation of information about genes, proteins and other cellular constituents is still clearly insufficient for a true comprehension of how organisms work. We have identified most genes and proteins in simple organisms, such as yeast and *Escherichia coli*, but we are still unable to make reliable predictions as to how these organisms would respond in an untested environment. We have a list of the three billion base pairs that define what it means to be human, but we are still far from understanding health and disease. This lack of true comprehension has direct and costly consequences for drug development. Pharmaceutical companies have an inventory of hundreds of thousands of compounds, yet the typical success rate for drug discovery is only ~4%, at a cost of several hundred million dollars for executing the necessary preclinical and clinical trials and bringing the drug to market [1].

If the reductionist paradigm is insufficient, what is impeding true understanding? The missing piece of the puzzle is often the reassembly of the analyzed, isolated component parts into a functioning conceptual entity [2,3]. This reconstruction is by no means trivial, because the biological and metabolic systems governing the effects of a drug are notoriously complex and the study of isolated metabolites and enzymes alone seldom reveals

1359-6446/02/$ – see front matter  PII: S1359-6446(00)02280-8 

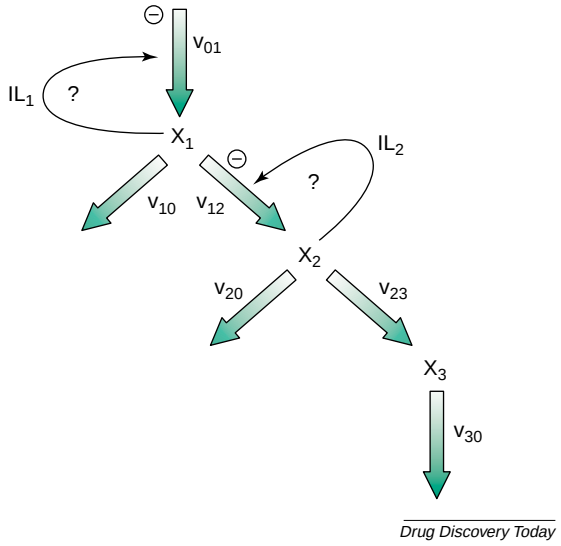


**Figure 1.** A generic pathway with two branch points (at $X_1$ and $X_2$) and two potential feedback inhibition loops, $IL_1$ and $IL_2$ (thin arrows), which might be present or absent, as indicated by question marks. The goal is to maximize output through the flux $v_{30}$. Intuition might suggest that, for example, it is immaterial whether one doubles $v_{12}$ or halves $v_{10}$. Indeed, in the absence of inhibition, this conclusion is correct. However, rigorous analysis with a canonical model of the pathway shows that this is not true if $IL_1$ is present, and that the optimal strategy for maximizing output depends on the presence or absence of $IL_1$ and/or $IL_2$.

their multitudinous roles *in vivo*. As a non-pharmacological illustration, even a detailed, rigorous study of the gases hydrogen and oxygen would probably not lead the chemist to predict key properties of water, such as its wetness [4]. A cell is more than a collection of membranes, organelles and proteins, mixed with some DNA and RNA.

The challenge in dealing with complex systems is a result of synergistic properties, which do not exist in any of their constituents but only emerge in their intricate interrelationships. This implies that reductionism must be complemented with a synergistic systems approach of reconstruction. The emerging reconstructionist paradigm thus freely acknowledges that reductionist methods and investigations of detailed mechanistic process are crucially important, but posits that they need to be accompanied by mathematical concepts that are capable of capturing the essence of complex, integrated systems [2,3,5–8].

## Synergistic systems as emerging paradigm

To appreciate the challenge of reconstruction, it is necessary to study in greater detail what makes biomedical systems complex and synergistic. The first criterion is a large number of constituents. The human genome contains 20,000–40,000 genes and with modifications these ultimately lead to several hundred thousand different proteins and peptides. Accounting for sugars, lipids and other small molecules renders the total number of cellular constituents much larger. Dealing with large numbers of constituents poses a significant book-keeping problem.

The second key criterion of biological complexity is the typically rich network of interactions among the constituents. These interactions are numerous and have nonlinear characteristics that are difficult to handle with intuition alone. For instance, a doubled input does not necessarily lead to a doubled response, and even a small change in the value of some component might cause the system to respond in an entirely new fashion. For the 'normal' value of some parameter, the system operates at some homeostatic 'steady-state' point, but if some stimulus causes the value to increase above a certain threshold, the system begins to oscillate or ceases to function at all. The difference between normal tanning and sunburn is a simple illustration for such a threshold response. Nonlinearities make complex systems difficult to understand and predict in their responses because they defy our innate way of reasoning in terms of chains of causes and effects.

### *Complexity at the pathway level*

Systems with intrinsic features of complexity are found at all levels of biological organization. A simple illustration might suffice as indication of the challenges encountered in nonlinear systems. Consider a generic pathway with two sequential branch points and two potential inhibitory feedback loops; as a biological example, one might envision a section of a metabolic network (Fig. 1). Suppose the goal is to increase the output of $X_3$, and the available tools of alteration are manipulations of any of the branches $v_{10}$, $v_{12}$, $v_{20}$ and $v_{23}$. If the total alteration effort were to be the same, where should it be expended? Does it matter whether one increases $v_{12}$ or decreases $v_{10}$? Would it be more efficacious to increase $v_{12}$ four times or to double $v_{12}$ and $v_{23}$? Mathematical analysis reveals that the answer depends on the existence of one or both inhibitory loops $IL_1$ and/or $IL_2$ [9]. Without inhibition, it is largely immaterial whether the branch toward the desired direction (e.g. $v_{12}$) is increased by some factor or whether the undesired branch ($v_{10}$) is decreased by the same factor. However, if the inhibitory loop $IL_1$ is present, the two strategies lead to different results, with an increase in $v_{12}$ being more effective than a decrease in $v_{10}$. If both inhibitory loops are present, the best strategy is strong activation of the flux $v_{23}$. Such results are difficult to obtain with intuition. Some features of complex systems might even be counterintuitive.

For instance, in the pathway of Fig. 1, inhibition can actually increase the potential for enhancing the yield and output of $X_3$. This simple example demonstrates that apparently unassuming differences in the regulatory structure of a system require specific strategies for optimal manipulation. Intuitive reasoning alone is not sufficient to predict such strategies (or operating principles), even for moderately sized systems that are regulated by activators and inhibitors.

The analysis of complex synergistic systems necessitates mathematical approaches that are rich enough to capture the multitude of possible system responses, from saturation to oscillations and possibly even chaos, yet enable efficient evaluation and facilitate insight and explanation. These models must be positioned somewhere between unstructured models-of-data, such as regression models, and highly detailed, mechanistic models-of-processes, which tend to become complicated so rapidly that they might obscure global patterns or trends in responses [10].

### *Traditional genome and metabolic analysis*

Two types of models presently dominate the analysis of genomes and metabolic pathways. The default model for assessing genome-wide expression data is statistical clustering, which identifies genes that are coexpressed after a stimulus. The magnitude of upregulation in each cluster of genes indicates the degree of importance that these genes have in a given response. Clustering has great potential for the discovery of known and unknown genes involved in a variety of organismic reactions. For instance, comparing the expression of genes from normal cells and cancer cells might render it possible to identify pathways or processes that differentiate the two cell types.

The default model for a biochemical process is the Michaelis–Menten rate law with its generalizations toward several substrates, reversibility and different mechanisms of inhibition [11]. This rate law results from a mathematical formulation of the elemental chemical processes that occur in an enzyme-catalyzed reaction and implicitly assumes homogeneity of the medium in which the reaction takes place. Although uncounted experiments *in vitro* have lent support to this rate law, three arguments make its validity questionable for the analysis of larger pathways and systems *in vivo*. First, as Schulz pointed out [11]; 'the concept of an enzyme catalyzing a reaction in isolation runs counter to the purpose for which enzymes have been provided in nature. The purpose of an enzyme in nature is to catalyze a reaction in concert with the other enzymes in the metabolic pathway, and the purpose of a metabolic pathway is to catalyze a series of reactions in concert with the many other pathways with which it interacts… If coordination of the action of…enzymes becomes flawed, it is certain that the living organism is going to encounter serious difficulty.'

Second, cells are strikingly different from a homogeneous mix in which substrates and enzymes exist. They are densely packed with organelles and metabolites [12] and many reactions take place in dimensionally restricted spaces, such as on membranes or within channels or small compartments. This dimensional restriction has an important impact on the mathematical concepts that should govern the kinetics of enzyme-catalyzed reactions. Although some theoretical work has addressed this topic [13,14], experimental quantification is still lacking.

The third issue with the use of the Michaelis–Menten rate law *in vivo* is its mathematical form. For a single reaction, the rational function describing the relationship between substrate concentration and flux is simple but as soon as many reactions are involved and the system is regulated by modulators, the mathematics becomes cumbersome [15]. One could argue that increased computing power would more than compensate for the greater complexity of larger systems. Indeed, more than 30 years ago, Garfinkel [16] and others constructed computer models in which the enzyme-catalyzed reactions were represented by rate laws in the tradition of Michaelis and Menten. However, Garfinkel's approach encounters two serious problems. One is the identification of parameter values, and the second is the analysis of these models and the interpretation of their results. Heinrich and Rapoport [17] described these problems in the following way: 'First, from the computer output it appears difficult to differentiate between important and unimportant effects, enzymes, metabolites, etc. Second, it is difficult to see how some effects are brought about. Third, such a computation is often impracticable for experimentalists. Fourth, many *ad hoc* assumptions are even now necessary for the mechanisms of several of the constituent enzymes of a chain. The strong dependence of the model of an enzymatic chain on the detailed mechanisms of single enzymes is unfavorable.'

Summarizing these types of considerations, Savageau [2] postulated an agenda for reconstruction in biology: 'The reductionist paradigm itself is inherently unable to deal with reconstructionist issues. We need a radically different but complementary approach that is able to elucidate quantitative and qualitative features of complex integrated systems…This will require adoption of new perspectives, development of new experimental methodologies for characterizing systemic behavior *in situ*, and an investment in 'bilingual' education [in biology and mathematics] for the next generation of biologists.'

## Canonical modeling

Assuming that we really target a true biomathematical approach, what types of models would be most useful and should be taught? At a superficial level, the answer is easy. In the context of drug discovery, we are looking for mathematical models of disease-related phenomena and these models should be valid, yet convenient for analysis and manipulation. Equipped with such models, we would be able to screen hypotheses and test ideas on the computer, which is faster and cheaper than a pharmacological study.

Although the goals are clear, the difficulty is that no single model entirely satisfies all items on our wish list. All models are compromises between their validity and applicability on one hand, and abstraction, simplification and mathematical feasibility on the other. Validity over a large spectrum of scenarios requires a high degree of complexity and the involvement of as many known factors and variables as possible. Simplicity of analysis requires just the opposite. Every biomathematical modeler has to struggle with this tension and there is no unique, generally justifiable solution.

### *Biochemical Systems Theory*

In the area of metabolic pathways and gene regulatory networks, one modeling framework has emerged over the past 30 years as being particularly useful. It is called canonical modeling [18] and is based on the rigorous mathematical foundation of Biochemical Systems Theory (BST) [19]. BST has been the subject of several hundred papers, chapters and other treatises, including a number of books and extensive reviews [3,8,15,20–22]. These publications have derived and analyzed the mathematics behind this formalism *in extenso*. Suffice it to say, the formalism is based on rigorous theorems of applied and numerical mathematics. Instead of reviewing these purely mathematical aspects (see Box 1 for a brief summary), it might be more beneficial here to talk about the two crucial features on which BST is based.

The first is a general tenet of systems analysis: one has to understand the dynamics of the components (i.e. the variables) of a system to understand the system itself. In the context of metabolic pathways, the variables are typically metabolites, enzymes or modulators, but could also include experimental conditions, such as temperature and pH. In physiologically based pharmacokinetics, each variable might represent the amount of a drug in a particular organ or the bloodstream, and in the analysis of gene regulatory networks, variables might represent transcription factors, repressors or specific gene products. The dynamics in each variable is described as the change in its value over time and this change is governed by the difference of all influxes and effluxes.

**Box 1. Construction of a canonical model [8]**

(1) Identify components to be included in the model.
(2) If they change over time, assign to them variable names $X_i$. If not, the components might still be variables by name or they might be absorbed in some of the model parameters.
(3) Identify the flow of material between variables.
(4) Identify regulatory signals, such as feedbacks.
(5) Create a diagram, as exemplified in Fig. 1.
(6) For each variable $X_i$ that changes over time, define an equation that relates its change over time to influxes and effluxes:

$$\text{Change in } X_i = \text{Fluxes into } X_i - \text{Fluxes out of } X_i$$

The change is equivalent to the derivative of the variable $X_i$, with respect to time: $dX_i/dt$
(7) Collect all influxes to $X_i$ in one function $F_i^{in}$ and all effluxes from $X_i$ in a second function $F_i^{out}$.
(8) Approximate $F_i^{in}$ and $F_i^{out}$ with power-law functions. If a flux F depends on only one variable X, the approximation is $F(X) \approx \alpha X^g$. If n variables govern the process, the analogous form is:

$$F(X_1, X_2, \ldots X_n) \approx \alpha X_1^{g1} X_2^{g2} \ldots X_n^{gn}.$$

The result of substituting these approximations in the equation of change (under 6) is always a canonical S-system equation of the form:

$$dX_i/dt = \alpha_i X_1^{gi1} X_2^{gi2} \ldots X_n^{gin} - \beta_i X_1^{hi1} X_2^{hi2} \ldots X_n^{hin}$$

(9) Estimate numerical values for all parameters from measurements or literature information. Each kinetic order $g_{ij}$ (or $h_{ij}$) uniquely represents the effect of $X_j$ on the flux $F_i^{in}$ or $F_i^{out}$, respectively. The parameters $\alpha_i$ and $\beta_i$ are rate constants.
(10) Analyze dynamics, steady states, robustness and responses under different scenarios, such as health, disease and drug treatment.

The second feature of BST is that all fluxes in the mass-balance system are approximated by so-called power-law functions (Box 1), the validity of which is mathematically justified by an old and famous theorem of Brook Taylor (1685–1731). The structural homogeneity and simplicity of power-laws might give the impression that models based exclusively on these functions are restricted in scope. However, this is a faulty conclusion: it has been proven with mathematical rigor that such models can represent virtually any set of smooth nonlinear phenomena [18,23]. Power-law functions have several properties that make them suitable for biological modeling purposes. In addition to the mathematical richness in structure, the list includes the following:

(1) Power-law functions are direct generalizations of elemental chemical reactions [15]. For instance, the formation of product in the elemental reaction $2A + B \Rightarrow C$ is traditionally represented by the (power-law) rate law; $v = k \cdot A^2 \cdot B$, where k is the rate constant and the powers of A and B (2 and 1, respectively) are kinetic orders. In canonical models, the rates have the same structure but the kinetic orders might be non-integer.

(2) Intracellular reactions apparently follow power-law kinetics [24]. This was shown with direct measurements *in vivo* and provides further credence to the validity of these functions.

(3) Power-law functions often capture metabolic and genetic observations over several orders of magnitude. For instance, Savageau [15] showed that the induction characteristic of the arabinose operon follows a power function over a 100-fold range of variation in the inducer arabinose.

(4) If taken to the molecular level, Michaelis–Menten rate laws are the approximation result of a power-law model [25]. This is because the Michaelis–Menten mechanism consists of elemental chemical kinetics, which themselves are described by power functions [see comment (1)].

(5) Hundreds of observations document that the growth rates of different tissues and organs in developing organisms are not independent of each other but closely follow the so-called law of allometry [26]. The same is true about a large variety of metabolic processes occurring within the same organism. Allometry means mathematically that a graph of one variable against the other is linear, if the variables are plotted as logarithms. This is exactly equivalent with a power-law model in linear (Cartesian) coordinates. Canonical models of the type shown in Box 1, called S-systems, satisfy the law of allometry, but most other mathematical models do not [8,27].

(6) The canonical model in Box 1 satisfies all of the previous features. In addition, the equations characterizing its steady state are linear [19]. A steady state is a condition of the system where all influxes exactly balance the effluxes so that all variables are constant in value over time, even though material flows through the system. Many systems in nature and in the laboratory reside in this state. The linearity of the steady-state equations is of great importance because it simplifies algebraic and numerical analyses [15] and permits efficient optimization of networks of pathways in biotechnology and metabolic engineering [28–30], which otherwise would be complicated and expensive in terms of computation time.

(7) Beyond steady-state characteristics, the homogeneous power-law structure of canonical models permits efficient methods of dynamical analysis [15,31] and numerical root finding [32].

(8) Power-law models are the basis for the Method of Controlled Mathematical Comparisons [33–35]. This method permits the specific assessment of particular model structures in alternative systems that are otherwise equivalent. For instance, it enables us to pinpoint the specific role of a regulatory signal: What are the advantages of having this signal? How would a system without it be disadvantaged? Similarly, the method enables objective assessments of the operation of a system under different environmental conditions. A typical question here is: Why is this gene upregulated, but others of the same pathway are not? [9,36].

(9) Although the prime focus of BST has been on metabolic and genetic systems, canonical models have been applied in a variety of areas. Examples include topics as diverse as microbial drug resistance [37], immunology [34], infectious diseases [38], fishery management [39] and forestry [40,41]. The success of these applications lends further support to the use of power-law functions as valid representations of complex biological systems.

(10) The power-law structure of canonical models translates directly from low to high dimensions (Box 1) and this has rendered it possible to analyze effectively biomedical phenomena with dozens of variables [25,42].

### *Applications of BST*

So, what can actually be done with canonical models? The spectrum of answers ranges from purely academic to practical and applied. At the academic end of the spectrum, the models have led to insights about the generic structure and organization of biochemical and genetic systems. For example, demand theory within BST correctly predicted the regulatory mode of dozens of genes in bacteria. Formerly, it had been assumed that it was an accident of nature whether a gene was normally switched off and only turned on upon demand, or whether it was normally switched on and repressed when necessary. However, using the canonical modeling approach described here, Savageau [43] was able to provide objective rationale for each type of regulation based on the organism's needs in its typical environment. Subsequently, numerous examples – and no exceptions – were found in support of the model predictions. In a similar vein, the rationale for different observed implementations of feedback loops was explained with methods of BST [15,35].

As a practical application, BST and the custom-tailored PLAS© (http://correio.cc.fc.ul.pt/~aenf/plas.html) for canonical

models facilitate simulations of complex biological scenarios. For instance, Curto and collaborators [44,45] developed a comprehensive model of human purine metabolism. This model could be used for screening new drugs. Suppose the goal were to develop a drug that would lower uric acid output. Almost three dozen enzymes are involved in purine metabolism and each could be a potential drug target. Clearly, testing all possible drug effects in the lab and in clinical trials would be expensive and it would be much cheaper to test a hypothetical drug with the mathematical model, in the following fashion. Because each enzymatic step has a unique representation in the canonical model, the parameters characterizing its kinetic features are easily identified. They typically consist of just one kinetic order and one rate constant (Box 1). The direct effect of the hypothetical drug on the enzyme under investigation is implemented in the model through the numerical alteration of the corresponding enzyme activity, possibly in a time-dependent fashion. For example, if the hypothetical drug is expected to reduce enzyme activity by 60%, this quantity is implemented in the corresponding enzymatic step of the model. Algebraic and dynamical analyses then show the predicted effects. They not only include the feature of prime interest, namely here uric acid output, but also altered levels of other metabolite levels, the relevance of which has to be evaluated by biochemists and clinicians. The model also reveals the dynamic transients in metabolites between the original, diseased state and the state upon administration of the drug. For instance, it could happen that, although the drug would ultimately lead to a lower uric acid level, some other metabolites would reach unacceptably high or dangerously low concentrations. The model could also reveal that even significant changes in a particular step would be ineffectual in lowering uric acid, and this result would remove the hypothetical drug from the list of top candidates. Through extensive simulations of this kind, a canonical pathway model can be used to explore single bolus experiments as well as different drug dosage and timing regimens [8]. This application is conceptually comparable with that of physiologically based pharmacokinetic models (D.B. Janszen, PhD Thesis, Medical University of South Carolina, 1992).

### *Metabolic engineering*

Another example for a practical application of BST falls in the area of metabolic engineering, whose general goal it is to redistribute fluxes within a metabolic network toward a desired goal. For example, we analyzed the goal of increasing citric acid production in the mold *Aspergillus niger* [30], an organism that has been used for this purpose for almost 100 years. Although the goal of the alteration is clear, numerous precautions are necessary to ensure that by-products do not accumulate to become toxic or that the organism suffers from 'metabolic burden', which tends to reduce viability and productivity [22,46]. Canonical models are uniquely beneficial for this kind of constrained optimization because they validly represent the metabolic and regulatory structures of networks of pathways, yet permit methods of linear programming [22,28]. In contrast to nonlinear optimization, which even for moderately sized pathways becomes a costly task, linear programming is easily executed with standard software for biochemical systems of hundreds of variables and constraints. Recently, Alvarez–Vasquez *et al.* (unpublished data) optimized the production of the medically relevant trimethylated amino acid, L-carnitine, in *E. coli* and an experimental laboratory [47] confirmed the theoretical predictions.

## Conclusions

In a position paper on genome sequencing, Edgar *et al.* [48] stated that the Human Genome Project was a 'tour de force' and that it was 'not at all clear that knowledge of the nucleotide sequence of the human genome [would], initially, provide deep insights into the physical nature of man.' Sixteen years later, the 2001 brochure *Genomes to Life* (http://DOEGenomestoLife.org), issued by the US Department of Energy, states as the 'next logical step [beyond international genome sequencing projects]…the quest to understand the composition and function of the biochemical networks and pathways that carry out the essential processes of living organisms.'

This next logical step will require four components, which are all within reach. First, the accuracy of gene expression profiles must be increased, so that it can be decided if a 1.5-fold increase in upregulation is real or noise.

Second, methods must be developed and improved for obtaining precise and encompassing measurements of protein and small molecule concentrations. One avenue of research is the construction of high-throughput proteome microarrays [49], which show profiles of all proteins that are available in a cell under defined conditions. These arrays are conceptually similar to DNA arrays but show directly current protein levels, instead of gene expression that ultimately might or might not translate into functional protein. Thus, if protein responses following a stimulus can be 'profiled' with reliability, questions concerning selective transcription of genes and post-translational modifications of their immediate products become less important for assessments of organismic responses at the metabolic level. In addition to the characterization of proteomes, it is becoming feasible to establish metabolite profiles with MS or NMR methods. Such profiles contain simultaneous measurements

of the concentrations of hundreds of molecules in various size ranges, from small molecules to lipids to proteins and nucleic acids. Obtained as time series, these profiles will be invaluable sources for identifying and quantifying the structure and properties of metabolic pathways *in vivo*.

Third, intensive research will be needed to decipher the exceedingly complex mechanisms of signal transduction [50–52]. If we truly understand the strategies and mechanisms with which a cell or organism responds to outside signals, it might become possible to use signaling molecules as specific targets that would require only minute amounts of drug to be efficacious. In mammalian cells, the signaling mechanisms are, at present, simply overwhelming in their complexity and redundancy. Although it will take some time to grasp these systems, simpler, yet similar signaling pathways are operative in yeast and other microbes and offer some hope for insight that might eventually be applicable to signal transduction in humans [53].

Fourth, we need to hone our mathematical modeling skills to integrate the enormous amounts of diverse types of data that are being produced in molecular biology and genome research. One starting point is metabolic flux analysis, which has the advantage of being linear and easily scaled up to organism-wide networks [54–58]. However, the key property of linearity is intimately intertwined with the greatest limitation of this approach, namely that nonlinear kinetic and regulatory features cannot be included in a straightforward fashion. This can become problematic, as in the simple pathway example of Fig. 1, where the regulatory structure dictates optimal strategies for effective operation. It is here that metabolic modeling with methods of BST could make the greatest contributions by explaining gene expression and metabolic profiles at the biochemical and physiological levels [36]. In comparison to purely stoichiometric approaches, canonical models require a lot of data, which are needed to decipher and quantify the kinetic and regulatory structure of metabolic pathways. In the past, these data were seldom available and a stoichiometric analysis was considered a valuable substitute. However, modern experimental methods of genomics, proteomics and metabolic profiling are rapidly changing the playing field, and comprehensive quality data will soon be available in great quantities for integrative analyses. This will enable detailed analyses not only of flux distributions but also of the more intricate strategies with which organisms control and regulate metabolism. Canonical models are good candidates for the execution of these types of analyses.

## Acknowledgements

I am grateful to Daniel Knapp and Ann Voit for constructive criticism. This work was supported in part by a Quantitative Systems Biotechnology grant (BES0120288) from the National Science Foundation.

## References

1 Bains, W. and Evans, C. (2001) The business of biotechnology. In *Basic Biotechnology.* (Ch. 12), (Ratledge, C. and Kristiansen, B., eds), Cambridge University Press
2 Savageau, M.A. (1991) The challenge of reconstruction. *New Biol.* 3, 101–102,
3 Savageau, M.A. (1992) Critique of the enzymologist's test tube. In *Fundamentals of Medical Cell Biology*. (Vol. 3A), (Bittar, E.E., ed.), pp. 45–108, JAI Press
4 Davidson, M. (1983) *Uncommon Sense.* J.P. Tarcher
5 Laszlo, E. (1972) *The Systems View of the World.* George Braziller
6 Yates, F.E. (1977) Explanation in science: is there a general theory of systems? *Am. J. Physiol.* 233, R169–R170
7 Service, R.F. (1999) Exploring the systems of life. *Science* 284, 80–83
8 Voit, E.O. (2000) *Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists.* xii + 532 pp., Cambridge University Press
9 Voit, E.O. Design and operation: keys to understanding biological systems. In *Function and Regulation of Cellular Systems: Experiments and Models.* Birkhäuser-Verlag (in press)
10 Voit, E.O. Models-of-data and models-of-processes in the post-genomic era. *Math. Biosci.* (in press)
11 Schulz, A. R. (1994) *Enzyme Kinetics. From Diastase to Multi-Enzyme Systems.* x + 246 pp., Cambridge University Press
12 Goodsell, D.S. (1993) *Machinery of Life.* Springer-Verlag
13 Kopelman, R. (1991) Reaction kinetics in restricted spaces. *Israel J. Chem.* 31, 147–157
14 Savageau, M.A. (1998) Development of fractal kinetic theory for enzyme-catalyzed reactions and implications for the design of biochemical pathways. *BioSystems* 47, 9–36
15 Savageau, M.A. (1976) *Biochemical Systems Analysis. A Study of Function and Design in Molecular Biology.* xvii + 379 pp., Addison–Wesley, Reading, MA, USA
16 Garfinkel, D. (1968) The role of computer simulation in biochemistry. *Comp. Biomed. Res.* 2, 31–44
17 Heinrich, R. and Rapoport, T.A. (1974) A linear steady-state treatment of enzymatic chains: general properties, control and effector strength. *Eur. J. Biochem.* 42, 89–95
18 Voit, E.O. (ed.) (1991) *Canonical Nonlinear Modeling. S-System Approach to Understanding Complexity.* xi + 365 pp., Van Nostrand Reinhold
19 Savageau, M.A. (1969) Biochemical systems analysis. *J. Theor. Biol.* 25, 365–379
20 Savageau, M.A. (1995) Enzyme kinetics *in vitro* and *in vivo*: Michaelis–Menten revisited. In *Principles of Medical Biology*. (Vol. 4), (Bittar, E.E., ed.), pp. 93–146, JAI Press
21 Voit, E.O. (2000) Canonical modeling: a review of concepts with emphasis on environmental health. *Environmental Health Perspectives*, 108, (Suppl. 5), *Mathematical Modeling in Environmental Health Studies*, 895–909
22 Torres, N.V. and Voit, E.O. *Pathway Analysis and Optimization in Metabolic Engineering.* Cambridge University Press (in press)
23 Savageau, M.A. and Voit, E.O. (1987) Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Math. Biosci.* 87, 83–115
24 Kohen, E. *et al.* (1974) Studies on metabolic events in localized compartments of the living cell by rapid microspectro-fluorometry. *Adv. Biol. Med. Phys.* 15, 271–297
25 Shiraishi, F. and Savageau, M.A. (1992) The tricarboxylic acid cycle in *Dictyostelium discoideum*. I. Formulation of alternative kinetic representations. *J. Biol. Chem.* 267, 22912–22918
26 Savageau, M.A. (1979) Allometric morphogenesis of complex systems: derivation of the basic equations from first principles. *Proc. Nat. Acad. Sci. U. S. A.* 76, 6023–6025

**27** Voit, E.O. (1992) Selecting a model for integrated biomedical systems. In *Biomedical Modeling and Simulation*. (Eisenfeld, J. *et al.*, eds), Elsevier Science Publishers

**28** Voit, E.O. (1992) Optimization in integrated biochemical systems. *Biotechnol. Bioengin.* 40, 572–582

**29** Hatzimanikatis, V. *et al.* (1996) Optimization of regulatory architectures in metabolic reaction networks. *Biotechnol. Bioeng.* 52, 485–500

**30** Torres, N.V. *et al.* (1996) Optimization of nonlinear biotechnological processes with linear programming. Application to citric acid production in *Aspergillus niger*. *Biotechnol. Bioeng.* 49, 247–258

**31** Irvine, D.H. and Savageau, M.A. (1990) Efficient solution of nonlinear ordinary differential equations expressed in S-system canonical form. *SIAM J. Numer. Anal.* 27, 704–735

**32** Savageau, M.A. (1993) Finding multiple roots of nonlinear algebraic equations using S-system methodology. *Appl. Math. Comput.* 55, 187–199

**33** Savageau, M.A. (1985) A theory of alternative designs for biochemical control systems. *Biomed. Biochim. Acta* 44, 875–880

**34** Irvine, D.H. and Savageau, M.A. (2000) Network regulation of the immune response. *J. Immunol.* 134, 2100–2130

**35** Alves, R. and Savageau, M. (2000) Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics* 16, 786–798

**36** Voit, E.O. and Radivoyevitch, T. (2000) Biochemical systems analysis of genome-wide expression data. *Bioinformatics* 16, 1023–1037

**37** Berg, P.H. *et al.* (1996) A pharmacodynamic model for the action of the antibiotic Imipenem on *Pseudomonas in vitro*. *Bull. Math. Biol.* 58, 923–938

**38** Voit, E.O. (1990) S-system analysis of endemic infections. *Comput. Math. Appl.* 20, 161–173

**39** Chaudhuri, K. and Johnson, T. (1990) Bioeconomic dynamics of a fishery modeled as an S-system. *Math. Biosci.* 99, 231–249

**40** Torres, N.V. (1996) S-system modelling approach to ecosystem: application to a study of magnesium flow in a tropical forest. *Ecol. Model.* 89, 109–120

**41** Voit, E.O. and Sands, P.J. (1996) Modeling forest growth. *Ecol. Model.* 86, 51–89

**42** Ni, T-C. and Savageau, M.A. (1996) Application of biochemical systems theory to metabolism in human red blood cells. *J. Biol. Chem.* 271, 7927–7941

**43** Savageau, M.A. (1998) Demand theory for gene regulation. *Genetics* 149, 1665–1691

**44** Curto, R. *et al.* (1998) Mathematical models of purine metabolism in man. *Math. Biosc.* 151, 1–49

**45** Curto, R. *et al.* (1998) Analysis of abnormalities in purine metabolism leading to gout and to neurological dysfunctions in man. *Biochem. J.* 329, 477–487

**46** Bentley, W.E. *et al.* (1990) Plasmid-encoded protein: the principal factor in the 'metabolic burden' associated with recombinant bacteria. *Biotechnol. Bioeng.* 35, 668–681

**47** Cánovas, M. *et al.* (2002) Modelling of the biotransformation of crotonobetaine into L(-)-carnitine by *Escherichia coli* strains. *Biotechnol. Bioeng.* 77, 764–775

**48** Edgar, R.E. *et al.* (1985) A paper prepared for the Human Genome Institute Workshop, May 1985, Santa Cruz, CA, USA; cited in Weinstein, J. (1998) Fishing expeditions. *Science* 282, 627

**49** Zhu, H. *et al.* (2001) Global analysis of protein activities using proteome chips. *Science* 293, 2101–2105

**50** Weng, G. *et al.* (1999) Complexity in biochemical signaling systems. *Science* 284, 92–96

**51** Abrieu, A. *et al.* (2001) The interplay between cyclin-B-Cdc2 kinase (MPF) and MAP kinase during maturation of oocytes. *J. Cell Sci.* 114, 257–267

**52** Kyriakis, J.M. and Avruch, J. (2001) Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. *Physiol. Rev.* 81, 807–869

**53** Hannun, Y.A. *et al.* (2001) Enzymes of sphingolipid metabolism: from modular to integrative signaling. *Biochemistry* 40, 4893–4903

**54** Jorgensen, H. *et al.* (1995) Metabolic flux distribution in *Penicillium chrysogenum* during fed-batch cultivations. *Biotechnol. Bioeng.* 46, 117–131

**55** Pons, A. *et al.* (1996) Metabolic flux distribution in *Corynebacterium melassecola* ATCC17965 for various carbon sources. *Biotechnol. Bioeng.* 51, 177–189

**56** Edwards, J.S. and Palsson, B.Ø. (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics and capabilities. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5528–5533

**57** Schilling, C.H. *et al.* (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* 15, 296–303

**58** Schilling, C.H. *et al.* (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203, 229–248


## Free journals for developing countries

The WHO and six medical journal publishers have launched the Access to Research initiative, which enables ~70 developing countries to gain free access to biomedical literature through the Internet.

The science publishers, Blackwell, Elsevier Science, the Harcourt Worldwide STM group, Wolters Kluwer International Health and Science, Springer–Verlag and John Wiley, were approached by the WHO and the British Medical Journal in 2001. Initially, >1000 journals will be available for free or at significantly reduced prices to universities, medical schools, research and public institutions in developing countries. The second stage involves extending this initiative to institutions in other countries.

Gro Harlem Brundtland, director-general for the WHO, said that this initiative was 'perhaps the biggest step ever taken towards reducing the health information gap between rich and poor countries'.

See http://www.healthinternetwork.net for more information.