

SYNTAX: A Rule-Based Stochastic Simulation of the Time-Varying Concentrations of Positional Isotopomers of Metabolic Intermediates

DAVID M. COHEN AND RICHARD N. BERGMAN

University of Southern California School of Medicine, Department of Physiology and Biophysics, Metabolic Research Unit, Los Angeles, California 90033

Received November 10, 1993

We present a new approach to simulation of metabolic pathways. The syntactic approach combines a rule-based description of biochemical reactions with a stochastic model of chemical kinetics. Each syntactic rule describes the location in a product molecule to which a particular carbon of a reactant molecule will be transferred. Using specifically labeled substrates and known rates of chemical reactions, our simulation predicts the time-dependent changes in concentration of positional isotopomers of metabolic intermediates. (A positional isotopomer of a compound is an isomer that is determined by the positions of isotopes within the molecule, e.g., $[1,2-^{13}\text{C}]$ glucose and $[1,3,5-^{13}\text{C}]$ glucose.) For the simulation of the ^{13}C -positional isotopomers of the citric acid cycle in heart cells we require only 39 syntactic rules, compared to the 176 ordinary differential equations required by the traditional approach. Addition of chemical reactions to the simulation does not require changing the program code for the existing reactions. In comparison, addition of a reaction to a system described by differential equations requires altering the equations for all isotopomers of all reactants and products of the new chemical reaction. © 1994 Academic Press, Inc.

I. INTRODUCTION

Metabolic pathways are coordinated groups of biochemical reactions which convert one or more chemical compounds into other compounds. Owing to the importance of these pathways in providing energy for life, changes in the rates of specific metabolic pathways have been observed in cancer, diabetes mellitus, and other diseases. Thus quantitative assessments of changes in metabolic pathways in these diseases are important. Recently, a noninvasive technique using stable (i.e., nonradioactive) isotopes has become available for estimation of metabolic flux rates in human subjects. The use of nonradioactive isotopes virtually eliminates risk to the subjects of the spectroscopic measurements.

We focus on the use of a stable isotope of carbon (^{13}C) for use in nuclear magnetic resonance (NMR) spectroscopy. NMR spectroscopy detects the presence of ^{13}C atoms within molecules and also the presence of ^{13}C atoms at

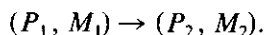
neighboring carbon positions. For example, consider the following two positional isotopomers¹ of acetic acid: [2-¹³C]acetic acid (chemical formula: HOOC-¹³CH₃) and [1,2-¹³C]acetic acid (chemical formula: HOO¹³C-¹³CH₃). The position in the NMR spectrum at which the signal from carbon C2 appears would contain a single peak in the spectrum of the former isotopomer and a doublet (two peaks) for the latter isotopomer. The latter isotopomer contains the second peak because of the interaction of the ¹³C atom at carbon C2 with the ¹³C atom at carbon C1. (Only neighboring carbon positions affect the NMR spectrum of carbon.) The abundance of different positional isotopomers determines the relative areas of the peaks of the NMR spectrum for any given compound (1). Although NMR spectroscopy can be used to measure the abundance of groups of positional isotopomers of a compound (2), current technology does not permit measurement of the concentration of individual isotopomers.

One of the challenges of the use of NMR spectroscopy is the proper design of accurate formulas and algorithms for estimation of metabolic flux rates. An important question in this process is "How can the formulas be tested?" Our answer to this question has been the construction of a computer simulation of the production of positional isotopomers that is independent of differential equations. Instead, we describe the changes to molecules in a structurally based simulation, in which populations of molecules are individually represented, including the ordered grouping of carbon atoms within each molecule. The result is a rule-based simulation, in which syntactic rules are used to describe the transfer of carbon atoms from reactants to products. In each transfer, owing to the literalness of the simulation, any ¹³C atoms in reactant molecules will be transferred to the "correct" positions in the product molecules (where the correctness of the transfer has been discovered by decades of chemical research).

II. BACKGROUND

A. Prior Work by the Authors

Our previous work on glycogen metabolism (3) and on the pattern of action of the enzyme heparin lyase (4) had demonstrated the advantage of a structurally based stochastic simulation. In our current simulation of the citric acid cycle (Fig. 1), we conceived of a set of syntactic rules as the repository of knowledge about the individual chemical reactions. We reduced the actions of a chemical reaction to a collection of rules describing the relocation of the atom at position P_1 in reactant molecule M_1 to position P_2 in product molecule M_2 (see Fig. 2):



(The form of these rules is intended to recall the form of production rules in

¹ A positional isotopomer of a compound is an isomer that is determined by the positions of isotopes within the molecule. For example, [1,2-¹³C]glucose, unlabeled glucose (all ¹²C atoms), and [1,3,6-¹³C]glucose are three different positional isotopomers of glucose.

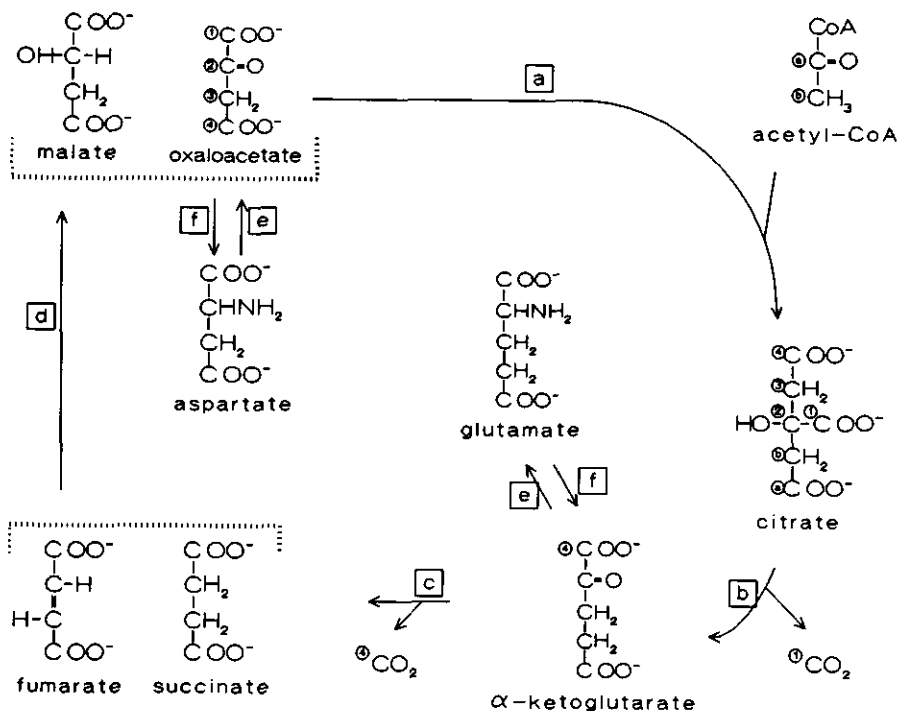


FIG. 1. Diagram of the citric acid cycle. Key to letters representing reactions catalyzed by the following enzymes: (a) citrate synthase; (b) aconitase and isocitrate dehydrogenase; (c) α -ketoglutarate dehydrogenase complex and succinyl CoA synthetase; (d) fumarase; (e) aspartate aminotransferase; and (f) aspartate aminotransferase (reverse). In the simulation, fumarate and succinate have been combined into a single pool; and similarly for malate and oxaloacetate. Carbons of oxaloacetate are numbered 1-4 and the carbons of the acetyl moiety of acetyl CoA are labeled a and b.

formal grammars (5).) We have published a complete set of syntactic rules for the simulation of the citric acid cycle and, where possible, have compared the results of our simulation to experimental measurements of others (6). In addition, the simulation has served as the arbiter of proposed formulas for estimation of flux through the citric acid cycle and the associated aminotransferase reactions (7).

The use of a population-based model, rather than a differential equations-based model, is a departure from the traditional technique for modeling metabolic pathways. Instead of the usual kinetic description of enzymatically catalyzed reactions, in which an algebraic variable represents each property of interest (8), the syntactic approach provides a miniature replica of the actual system, with each simulated molecule representing some micromoles of actual molecules. It also has the advantage of simplifying the description of the system. Instead of 176 coupled ordinary differential equations whose numerical solution provides the concentration of each isotopomer as a function of time, the syntactic approach requires only 39 syntactic rules. Use of syntactic rules to represent

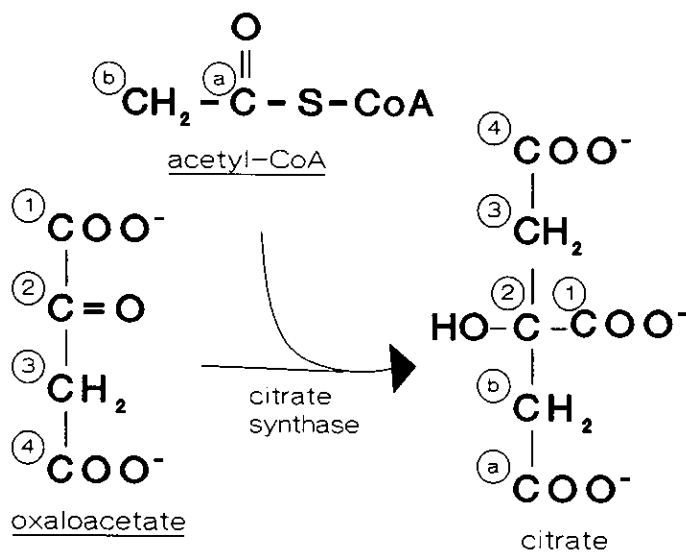


FIG. 2. Illustration of the syntactic rule which describes the transfer of atoms from reactants to products in the chemical reaction catalyzed by citrate synthase. The carbons at positions C1, C2, C3, and C4 of oxaloacetate are transferred to positions C6, C3, C2, and C1, respectively, of citrate. The carbons at positions C1 and C2 of the acetyl moiety of acetyl CoA are transferred to positions C5 and C4, respectively, of citrate.

the description of the orderly transfer of atoms from reactants to products, adjoined to this structurally based population model of metabolic pools, functionally decouples the differential equations. In other words, additions to the set of reactions simulated by the computer program do not require any changes to the description of the existing reactions. In the case of differential equations, this would not be the case.

B. Related Work

Prediction of positional isotopomers. Chance and co-workers proposed a model of the citric acid cycle from which Fig. 1 was obtained (9). They wrote 176 ordinary differential equations (1 per positional isotopomer of each metabolic intermediate). Measuring the mass of most of the metabolites as well as the NMR spectrum of glutamate over time, they estimated the rate of flux of the citric acid cycle and related aminotransferase reactions. From their estimations of the values of the parameters of their model, they simulated the changes in concentration of individual positional isotopomers within the citric acid cycle. Their computer simulation represents the traditional approach to the prediction of positional isotopomers over time.

Stochastic models of chemical kinetics. Bartholomay (10-12) proposed stochastic differential equations for chemical kinetics which are consistent in the mean with deterministic differential equations, i.e., whose solutions

have mean values equal to the deterministic solution. In other words, the rates of chemical reactions which are described by differential equations are the mean values of the rates predicted by the stochastic models of chemical kinetics. Gillespie (13) derived stochastic differential equations for individual chemical reactions as well as groups of simultaneous chemical reactions. He provides examples of oscillating chemical systems for which the stochastic formulation of the dynamics predicts the future behavior more accurately than does the analytic solution to a (nonstochastic) differential equation (13). In particular, the stochastic formulation excels in predicting the future behavior of systems possessing a stable but not asymptotically stable equilibrium (13, 14).

III. DESIGN CONSIDERATIONS

In the kinetic theory of chemical reaction rates, the rate of a chemical reaction is determined by the number of molecules that amass enough energy via collision to surpass the energy of activation (8, 15, 16). Although random motion is responsible for the movement of reactant molecules in the liquid or gas phase, the liquid phase greatly restricts free movement of reactants by trapping reactant molecules within "cages" of solvent molecules. The Michaelis–Menten model of enzyme-catalyzed chemical kinetics postulates that the dissociation of the enzyme–substrate complex is much slower than the rate of conversion of the complex to product molecule(s). Therefore the occurrence of enzyme-catalyzed chemical reactions may be viewed as a process of random collisions of enzyme and substrate which immediately result in the transmutation of substrate molecules to product molecules. Hence stochastic or probabilistic models are appropriate descriptions of the molecular causes of chemical reactions in liquid as well as gaseous phases.

A probabilistic model of biochemical kinetics needs to capture the Markov property; i.e., that the future behavior of the system is completely determined by its current state (10, 17). Rates of reaction should be determined by the concentrations of enzyme and substrate alone, and not on the history of the chemical reactions. The Poisson process is a stochastic process that possesses the Markov property and that may be described as a counting of events, in which the number of events in the time interval $[s, t]$ is a random variable. The axioms of a Poisson process are the following (18, 19): (i) the number of events in disjoint time intervals are independent; (ii) the probability of an event occurring in a time interval of duration Δt is proportional to the length of that time interval, for Δt sufficiently small; and (iii) the probability of two or more events occurring during a time interval of duration Δt is zero for Δt sufficiently small. The independence of the number of events in disjoint intervals in Axiom (i) means that the number of events in an interval does not depend on the number of events in any other (nonoverlapping) interval. In other words, the occurrence of a chemical interaction does not directly affect the occurrence of subsequent interactions. Thus, the efficiency of the enzyme does not change over time (it does not become tired or worn out), in agreement with current

thinking about enzyme-catalyzed reactions (8). Axiom (ii) allows the introduction of a parameter λ , the rate of the process, which is analogous to the rate constant of deterministic models. Bartholomay demonstrated the agreement of this axiom with the physics of molecular collisions (11). Axiom (iii) asserts that successive events are separable in time and that the entire set of chemical reactions does not cluster in a (mathematically) pathological manner about a single time point. This last axiom would exclude, for example, a cluster of events (a chain reaction) which occurred with increasing rapidity over time.

Symmetry of a molecular species plays an important role in determining the isotopomers of the metabolic pools (8). There are two symmetric molecules (succinate and fumarate), for which positional isotopomers that are mirror images of one another cannot be distinguished chemically (they are "achiral"). For example, [2-¹³C]fumarate cannot be distinguished chemically from [3-¹³C]fumarate. Therefore, a molecule of [3-¹³C] α -ketoglutarate would produce [2-¹³C]succinate (same as [3-¹³C]succinate), which would produce [2-¹³C]fumarate (same as [3-¹³C]fumarate) which would produce either [2-¹³C]malate or [3-¹³C]malate, with equal probability. Citrate is a molecule which appears symmetric but whose orientation can be distinguished functionally by enzymes and substrates (it is "prochiral"). Hence [1-¹³C]citrate can be distinguished chemically from [5-¹³C]citrate (8). The simulation carefully treats the labeling of symmetric molecules. Whenever a symmetric molecule is selected as a reactant, the orientation of the numbering of the carbons of that molecule is assigned probabilistically.

The stochastic simulation technique provides an exact solution of the equations, without the use of a discrete approximation (Δt) of an infinitesimal change in time (dt) typically used in the numerical solution of differential equations (13). Because the reactions selected to occur within a discrete time step are assumed to be simultaneous, the order in which their simulation occurs is randomized to avoid bias in the result. Especially with relatively small numbers of simulated molecules, the product of one reaction may be the reactant in another reaction during a single time step. Although this occurs with probability inversely proportional to the number of simulated molecules, the large number of time steps in our simulations (more than 1.5 million time steps) magnifies this likelihood. Since the state of the biochemical simulation is determined by the molecules within each metabolic pool, the effect of such bias on the outcome will not be automatically corrected over time. Therefore, the relatively small number of simulated molecules (between 100 and 1000 per pool) may be sensitive to the lack of randomization of reactions. For these reasons, it is important to simulate the stochastic processes exactly, without making the assumption that all reactions occurring within a single discrete time step are simultaneous.

IV. SYSTEM DESCRIPTION

The program begins with the reading of a data file and the initialization of the simulation program (Fig. 3). The initial populations of molecular pools are created and data structures describing the actions of reactions on the atoms of

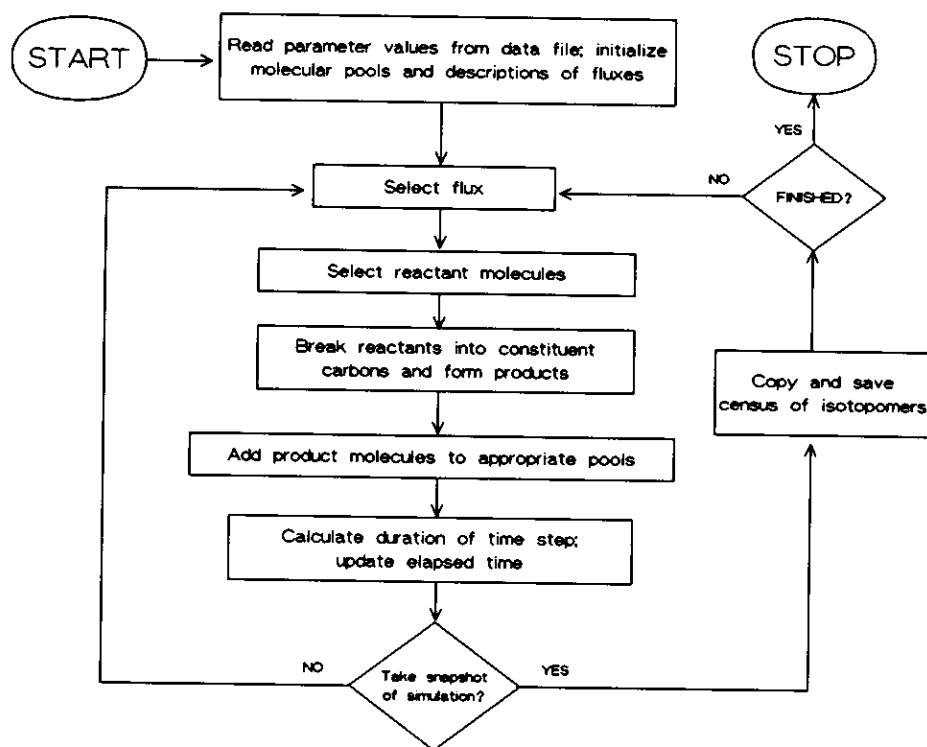


FIG. 3. Flowchart of the top-level of organization of the simulation program.

the reactant molecules are constructed. The data structure for a given reaction includes a linked list for the reactants and a linked list for the products. The paths from reactant chemical species to product species contain intermediate nodes which specify the placement of atoms of the reactant into positions of the product molecule (see Fig. 4).

After initialization, the program iterates the simulation of individual chemical reactions. As described in the Appendix, the identity of the reaction that occurs next is determined probabilistically. The probability of selecting a given reaction is proportional to its instantaneous rate, relative to the other reactions. The unit interval (0, 1) is partitioned into nonoverlapping segments each length of which is proportional to the current rate of one of the reactions. The algorithm for selecting a reaction (Fig. 5) requires computation of a single random number r which is then used to select an interval and corresponding reaction. (See the Appendix for details of the random number generator.) Having selected a reaction, we next select reactant molecules, decompose them into their atoms (we focus on carbon atoms only, for the present), and then recombine the atoms in accord with the actions of the given biochemical reaction (Fig. 3). The process of selecting a reactant molecule from a given metabolic pool requires the random selection from among all of the simulated molecules of that pool, which differ from

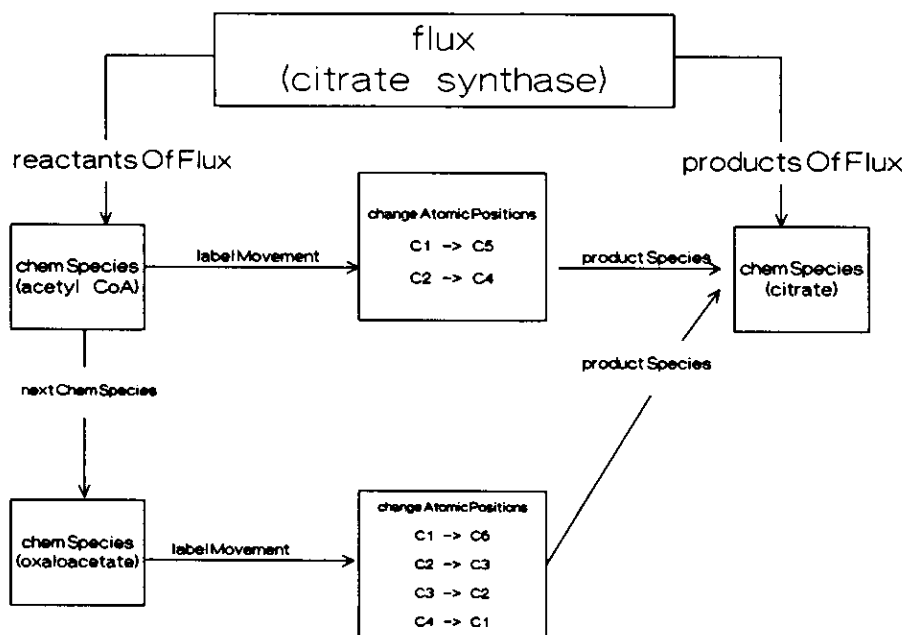


FIG. 4. Data structure for describing the reaction catalyzed by citrate synthase. One record per reactant and product molecule is maintained, with separate lists for reactants and products. The record labeled "changeAtomicPositions" contains information regarding the transfer of each atom of the reactants to the correct position in the product molecule. The carbons at positions 1 and 2 of acetyl CoA become the carbons at positions 5 and 4, respectively, of citrate. Similarly, the carbons at positions 1, 2, 3, and 4 of oxaloacetate are relocated to positions 6, 3, 2, and 1, respectively, of citrate. One such description is created for each reaction in the simulation.

one another only in their isotopomer type. If there are M molecules in a given pool then the unit interval $(0, 1)$ is partitioned into M equal and nonoverlapping subintervals. The probability of selecting a given subinterval is proportional to its length. A random number r , having a value in the unit interval, is used to select one of those subintervals. Therefore, the molecule corresponding to a given subinterval is chosen with probability $1/M$, as desired.²

After selecting the isotopomer type of the reactant molecule, a linked representation of that isotopomer is constructed and attached to the appropriate (chemSpecies) record in the linked list (reactantsOfFlux) of the appropriate reaction (see Fig. 4). When each of the reactant molecules has been selected, the description of the carbons of the reactant molecules is transferred to the

² Isotope discrimination in chemical reactions can easily be accommodated in the computer program, by assigning the probability of selecting a given molecule that depends on the masses of its atoms (i.e., depends on whether carbon is ¹²C or ¹³C). Similarly, other peculiar aspects of chemical reactions can be simulated, insofar as they affect the kinetics of the reactions.

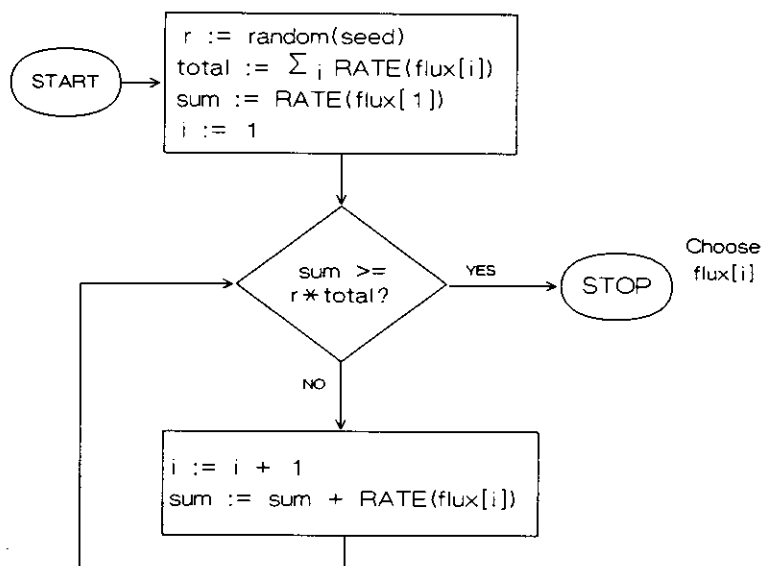


FIG. 5. Flow chart of algorithm for probabilistic selection of reaction. Each reaction (reaction[i]) has a probability of occurrence, proportional to its current rate (RATE(reaction[i])).

product molecules, guided by the information in the adjacent record (change AtomicPositions) and the link to the product molecule(s).³ After the destruction of the reactant molecules and the creation of the product molecule(s), the tally of isotopomers for the reactant and product pools is updated. The molecule(s) of product are then “decomposed” into atoms which are then saved (recycled) for use in constructing reactant molecules in the future. In this manner, the storage utilized by the computer program does not increase as the size of the populations of molecules changes.

Next the duration of the time step is computed and the elapsed simulated time is increased by the length of the time step. Since our interest does not lie in the distribution of elapsed times, we choose not to simulate the duration of each time step as a random variable. Instead, we let the value of the time step equal the mean or expected value of the actual time step. As shown in the Appendix, the expected value equals $1/\lambda_M$, where λ_M equals the sum of the rates of all simulated (concurrent) reactions. At regular intervals (in terms of the number of occurrences of chemical reactions) the state of the simulation is copied to a record of the historyList (illustrated in Fig. 6). Each history record contains the contemporaneous simulated elapsed time and the census of isotopomers in each metabolic pool. These records can then be used to create data files containing information on the changes over time of the concentration of isotopomers in each metabolic pool.

³ The atoms of carbon may differ from one another in molecular weight. Our current implementation allows three different isotopes of carbon: the most abundant isotope (carbon-12); a nonradioactive, stable isotope (carbon-13); and a radioactive isotope (carbon-14).

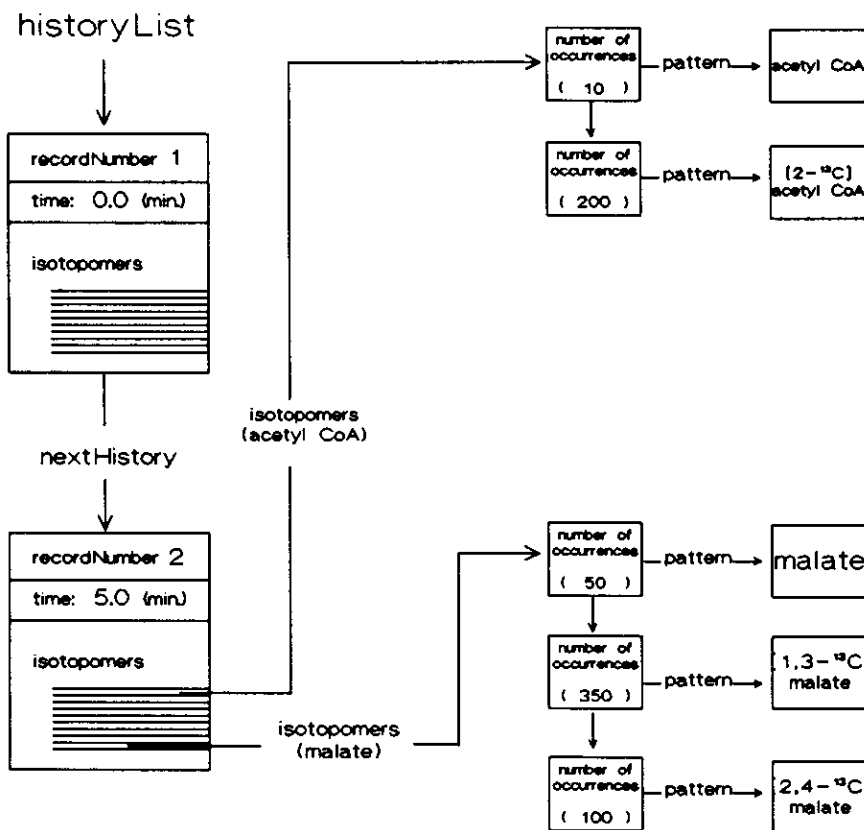


FIG. 6. Description of the historyList data structure, a linked list of records containing information saved at a particular point in the program. The "isotopomers" field contains pointers to the census of positional isotopomers of each metabolic pool. The number of simulated molecules of two isotopomers of acetyl CoA and three isotopomers of malate are shown.

A direct simulation, in which each molecule possesses a distinct memory location, is not economical of computer storage. Using a 1-1 mapping of molecules to computer storage locations increases the storage requirements of our program as the sizes of the metabolic pools increase. Furthermore, with large simulated populations, the direct simulation technique requires more execution time for management of dynamic data structures (20). We therefore prefer a representation in which just the *number* of simulated molecules of each isotopomer type (of each molecular pool) needs to be stored (Fig. 6). This allows the population of simulated molecules to grow in size and change its composition of isotopomers without either slowing down the simulation or increasing the amount of storage utilized. It does not affect the outcome of the simulation.

This compressed mode of representing the isotopomer concentrations of metabolic pools simplifies the taking of a "snapshot" of the simulation, i.e., recording the number of simulated molecules of each isotopomer type. In the case of a direct simulation of each molecule, it would be necessary to examine

each molecule to determine its isotopomer type in order to perform a census of the molecular populations. Obviously, the time required to perform this inventory of simulated molecules would increase linearly with the number of simulated molecules. However, using our compact representation, the time required is constant (and proportional to the number of possible isotopomers (176 isotopomer for the citric acid cycle)).

The simulation is currently running on a SPARCstation IPX with 16 Mbytes of random access memory (SUN Microsystems, Inc., El Segundo, CA). It requires 1.5 hr of cpu time to simulate 40 min of elapsed time, with a scaling factor of 4×10^{-4} (moles of actual molecules per simulated molecule). As expected, the execution time increases linearly with the scaling factor (14.5 hr of cpu time when the scaling factor is 4×10^{-5}).

V. STATUS REPORT

Chance and co-workers (9) published the only other prediction of time-dependent abundances of isotopomers of the citric acid cycle. They performed NMR measurements in perfused rat hearts subsequent to administration of glucose and [2- ^{13}C]acetate and fit their observations to a mathematical model described by differential equations. Their predictions of changes in isotopomers of glutamate and in the enrichments of individual carbons of glutamate (Figs. 4 and 6 of Ref. 9) are virtually identical to our predictions (Figs. 7a and 7b), even though their model of biochemical pathways and their computational methods are quite different from ours.

In addition to comparing the predictions of changes in the concentrations of isotopomers of intermediates of the citric acid cycle, we also compared the relative areas of the peaks of predicted NMR spectra at isotopic steady state with the findings of Malloy and co-workers (21). We were able to perform the comparison because Malloy and co-workers estimated the ratio of flux within the citric acid cycle to flux into pools of the citric acid cycle coming from pools not included in the citric acid cycle (so-called anaplerotic reactions). Furthermore, they estimated the labeling of the anaplerotic flux. Using the simulation, we duplicated the conditions for three different experiments. As we reported elsewhere (6), the agreement is excellent.

The isotope carbon-13 occurs naturally in approximately 1.1% of all carbons. We optionally include the natural occurrence of this isotope in the initial population of molecules as well as in the influx into any metabolic pool from pools not included in the simulation. Furthermore, we have added other metabolic pathways, including the pentose phosphate pathway, the glycolytic pathway, and the gluconeogenic pathway. We believe that one useful outcome of the syntactic simulation will be the prediction of isotopomers in more complex metabolic pathways and combinations of pathways.

VI. LESSONS LEARNED

A stochastic simulation of a syntactic model of biochemical pathways possesses several advantages over conventional approaches to simulation. The stochastic model of chemical kinetics eliminates the need to perform numerical

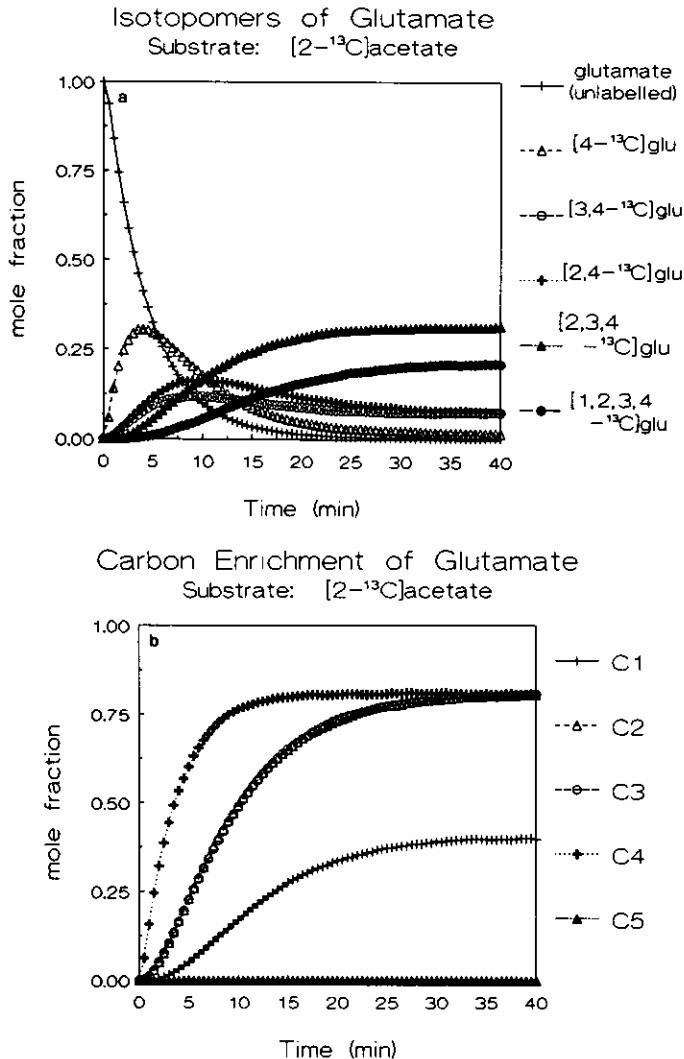


FIG. 7. Simulation of the citric acid cycle in heart cells, subsequent to administration of [2-¹³C]acetate instead of unlabeled acetate at 0 min. The pool sizes are at steady state, although the isotope distribution is not at steady state. (a) most abundant positional isotopomers of glutamate; (b) enrichment of individual carbons of glutamate (calculated from (a)).

integration of differential equations in order to predict the future behavior of the metabolic pools in a network of biochemical reactions. Although there is no inaccuracy in the stochastic approach (13), there is loss of precision due to the random fluctuations inherent in the chemical kinetics. We found that performing three repetitions with approximately 100–10,000 simulated molecules per pool reduced the standard error of the mean to less than 5%. Details

of individual reactions can be easily handled in a straightforward manner, because in principle, each atom of each molecule in the metabolic pools is literally represented in the computer memory. For example, metabolic channeling⁴ could easily be included, as well as isotope discrimination effects (23).

The advantages of a syntactic model are several. First, the generation of positional isotopomers of metabolic species follows naturally and immediately from the transfer of atoms of labeled reactant molecules to specific locations in the product molecules. Second, in the context of a computer simulation technique based on a Poisson process model of chemical kinetics (see above), the use of a syntactic representation avoids the need for integration of numerous differential equations (one per positional isotopomer). With this increased simplicity of the simulation program, details of individual reactions (such as asymmetries in reversible reactions or symmetries in molecular structure) can be accommodated easily. Third, the representation of biochemical reactions using the syntactic approach offers a tremendous advantage in software modularity over the traditional approach, which uses differential equations. Each addition of a chemical reaction to an existing syntactic simulation simply requires adding a small number of syntactic rules. In comparison, addition of a reaction to a system described by differential equations requires altering the equations for all isotopomers of all reactants and products of the new chemical reaction. Since the number of isotopomers of a single compound increases *exponentially* with the number of carbons in that compound, the number of equations that are affected can be considerable. Fourth, it is relatively easy to write an interface to the syntactic simulation that allows the user to describe the transfer of atoms in the chemical reactions of interest. Each run of the simulation would include the specification of the chemical reactions to be simulated. In this manner, a flexible computer simulation may be made available to the biochemist who does not possess expertise in computer programming.

VII. FUTURE PLANS

Investigations of the uniqueness of the syntactic approach need to continue. We intend to develop a stochastic optimization procedure that will estimate parameters of models of biochemical kinetics from observations of NMR spectra over time. This will enable us to apply our rule-based model to one of the basic problems of metabolic research *in vivo*: the study of changes to rates of metabolic flux in disease. We shall broaden our investigations to include multiple metabolic pathways in carbohydrate, lipid, and protein metabolism. The complexity of multiple pathways presents almost insurmountable obstacles to traditional approaches of analysis of metabolic pathways, especially during non-steady-state conditions. These theoretical investigations will have applica-

⁴ Metabolic channeling occurs when intermediates in a metabolic pathway are passed from one enzyme to another, without equilibrating with the solvent (usually water). The effect of channeling is that asymmetrical results are observed when symmetric molecules are involved (22).

tion to disease states for which the exact abnormality in metabolism has yet to be determined, such as non-insulin-dependent diabetes mellitus (24) and cancer (25).

Analysis of NMR spectra has already led to the discovery of new metabolic pathways (26). We will begin the analysis of the production of positional isotopomers in metabolic pathways using symbolic logic and techniques from artificial intelligence (27). The syntactic rules describing metabolic pathways lend themselves to the application of mathematical logic for the identification of positional isotopomers whose appearance is impossible without postulating new biochemical reactions. It may be possible for us to identify new biochemical pathways from carefully devised experiments with specifically labeled substrates and the logical analysis of isotopomers of known intermediates. In addition, we envision an expert system which could answer questions regarding the generation of positional isotopomers during administration of isotopically labeled compounds. Such a system might be used in the design of experiments utilizing NMR spectroscopy.

APPENDIX: POISSON PROCESSES AND CHEMICAL KINETICS

Duration of the time step. We describe how to compute the duration of each time step during the simulation of simultaneous chemical reactions. The three axioms of a Poisson process (see above) can be used to derive the probability that N_t events occur within a time interval of length t (18):

$$P\{N_t = j\} = (\lambda t)^j e^{-\lambda t} / j! \quad [1]$$

This is the Poisson probability density function with parameter λt , $t > 0$, where λ is the rate of the Poisson process. The expected number of events $E[N_t]$ occurring during a time interval of length t can be calculated from the following equation (19):

$$E[N_t] = \lambda t. \quad [2]$$

It can be demonstrated (19) that the time T between successive events described by a Poisson process (with parameter λ) is a random variable having an exponential probability density function $\lambda e^{-\lambda t}$, $t > 0$ and hence mean value $1/\lambda$ and variance $1/\lambda^2$. Formally, one may write

$$P\{T = x\} = \lambda e^{-\lambda x}. \quad [3]$$

The choice of the length of the time interval is made by sampling from the distribution of T , the time between successive events of a Poisson distribution.

Now consider \mathbf{N} simultaneous chemical reactions, each an independent Poisson process P_i with parameter λ_i and interarrival time T_i , $1 \leq i \leq \mathbf{N}$. Let $X_i(t)$ equal the number of occurrences of reaction i , $1 \leq i \leq \mathbf{N}$, in a given time interval $[s, s + t]$, for any $s > 0$, $t > 0$. Since the sum of a finite number of independent Poisson random variables is a Poisson random variable (19), the random variable $X_{\mathbf{M}}(t) = \sum_i X_i(t)$ has a Poisson distribution with mean $\lambda_{\mathbf{M}}$,

where $\lambda_M = \sum_i \lambda_i$. It can be shown that a single, composite Poisson process P_M with parameter λ_M satisfies the postulates of a Poisson process (19). As illustrated below, we can select time intervals short enough so that exactly one event of P_M occurs during each interval.

We illustrate the similarity of N simultaneous Poisson processes to a single, composite Poisson process with a small example. Consider three simultaneous Poisson processes A, B, and C, with $\lambda_A = 10 \text{ min}^{-1}$, $\lambda_B = 20 \text{ min}^{-1}$, and $\lambda_C = 40 \text{ min}^{-1}$. During any time interval of 1 min duration, the expected number of occurrences of processes A, B, and C is 10, 20, and 40, respectively. Thus the average number of occurrences of any of the three processes during the 1-min interval equals 70. The composite Poisson process has a parameter λ_M of $\lambda_A + \lambda_B + \lambda_C$ or 70 min^{-1} . The expected number of occurrences of the composite Poisson process during an interval of 1 min is 70.

Now, suppose we partition a 1-min interval into 70 equal, nonoverlapping subintervals and allow exactly one occurrence of either process A, B, or C in any subinterval. Then we should expect that, on the average, 10, 20, and 40 of the subintervals should represent the occurrence of processes A, B, and C, respectively. Now, if we probabilistically assigned processes A, B, and C to individual subintervals by letting the probability of these processes be $1/7$, $2/7$, and $4/7$, respectively, then we should expect the same distribution of processes within subintervals. Indeed this elementary, nonrigorous demonstration can be substantiated by mathematical analysis, based on the independence of nonoverlapping time intervals and the memoryless property of Poisson processes.

Selection of the chemical reaction. In simulating chemical reactions, we need to determine which of the N reactions occurs during a given time interval. For each interval, we need to select one of the N biochemical reactions (Poisson processes P_i , $1 \leq i \leq N$) as the unique event that occurs during the given interval. The probability of occurrence of process P_i can be calculated by defining the complementary Poisson process P_c , which includes all processes in P_M except process P_i . Let $X_c(t)$ denote the number of occurrences of process P_c in the time interval $[s, s + t]$, for any $s > 0$, $t > 0$, where

$$X_c(t) = \sum_{j \neq i} X_j(t). \quad [4]$$

Then P_c and P_i are two independent Poisson processes whose interarrival times (T_c and T_i , respectively) are exponentially distributed with means $1/\lambda_c$ and $1/\lambda_i$, respectively. Owing to the memoryless property of Poisson processes, at any time t , T_c , and T_i are the expected times until the next occurrence of P_c and P_i , respectively. The probability that P_i occurs prior to P_c equals the probability that process P_i is the next process to occur among the N concurrent processes and equals (19)

$$\lambda_i/\lambda_M,$$

as shown by the following derivation:

$$\begin{aligned}
 P\{T_i < T_c\} &= \int_0^\infty P\{T_i < T_c \mid T_c = x\} P\{T_c = x\} dx \\
 &= \int_0^\infty P\{T_i < x\} \lambda_c e^{-\lambda_c x} dx \\
 &= \int_0^\infty (1 - e^{-\lambda_i x}) \lambda_c e^{-\lambda_c x} dx \\
 &= 1 - \lambda_c / (\lambda_i + \lambda_c) \\
 &= \lambda_i / \lambda_M.
 \end{aligned}$$

Thus at each time interval, the rates of the biochemical reactions determine the parameters of the corresponding Poisson processes, which are used to assign a probability to each chemical reaction. We assign to each Poisson process P_i a probability equal to λ_i/λ_M , and which is used to select probabilistically the biochemical reaction for the current time interval. We have tacitly assumed the stationarity of the Poisson processes, i.e., that the Poisson parameters λ_i did not change during each time step, because the occurrence of one reaction within a single time step will not significantly change the pool sizes (the number of simulated molecules per pool) or the rates of chemical reactions during that time step.

Random number generator. The selection of pseudorandom numbers is an important aspect of the simulation algorithm. We use a multiplicative linear congruential generator, with the parameter values recommended by Park and Miller (28–31). A single parameter value (seed) determines the sequence of random numbers completely and is therefore varied with each simulation run. The algorithm for computing a random number (28) has two steps: (i) compute a new value for the seed from its current value,

$$\text{seed} := (\text{seed} * 16807) \text{ MODULO } (2^{31} - 1),$$

and (2) compute the next random number from the new value of the seed,

$$\text{randomNumber} := \text{seed} / (2^{31} - 1).$$

The simulation program protects against integer overflow by avoiding the multiplication of 16807 and the seed (28). The computer program uses the constants 127773 ($2^{31} - 1 \text{ DIV } 16807$) and 2836 ($2^{31} - 1 \text{ MOD } 16807$) and integer arithmetic to prevent overflow in computers with sufficient word size to represent integers as large as $2^{31} - 1$. The functions DIV and MOD are the integer functions available in the Pascal programming language. The program fragment for computing seed is the following:

```

i1 := seed MOD 127773;
i2 := seed DIV 127773;
seed := 16807 * i1 - 2836 * i2;
IF seed <= 0 THEN seed := seed + 231 - 1;

```

ACKNOWLEDGMENTS

This work was supported by a grant from the USC Faculty Research and Innovation Fund (FRIF) and by the NIH (AM27619). D. M. Cohen was supported by training grant AG00093 from the N.I.A. and R. N. Bergman is supported by the USC Salerno Collegium.

REFERENCES

1. MALLOY, C. R., SHERRY, A. D., AND JEFFREY, F. M. H. Analysis of tricarboxylic acid cycle of heart using ^{13}C isotope isomers. *Am. J. Physiol.* **259**, H987–H995 (1990).
2. JEFFREY, F. M. H., RAJAGOPAL, A., MALLOY, C. R., AND SHERRY, A. D. ^{13}C -NMR: A simple yet comprehensive method for analysis of intermediary metabolism. *Trends Biochem. Sci.* **16**, 5–10 (1991).
3. YOUN, J. H. AND BERGMAN, R. N. Patterns of glycogen turnover in liver characterized by computer modeling. *Am. J. Physiol.* **253**, E360–E369 (1987).
4. COHEN, D. M., AND LINHARDT, R. J. Randomness in the heparin polymer: Computer simulations of alternative action patterns of heparin lyase. *Biopolymers* **30**, 733–741 (1990).
5. HOPCROFT, J. E., AND ULLMAN, J. D. "Introduction to Automata Theory, Languages and Computation." Addison–Wesley, Reading, MA, 1979.
6. COHEN, D. M., AND BERGMAN, R. N. Prediction of positional isotopomers of the citric acid cycle: The syntactic approach. *Am. J. Physiol.* **226**, E341–E350 (1994).
7. COHEN, D. M., AND BERGMAN, R. N. Estimation of the rate of flux of the citric acid cycle. Scientific Conference on the Application of Magnetic Resonance to the Cardiovascular System, American Heart Association, Dec. 15–18, 1993, Atlanta, Georgia. (Abstract)
8. STRYER, L. "Biochemistry," 3rd ed. Freeman, New York, 1988.
9. CHANCE, E. M., SEEHOLZER, S. H., KOBAYASHI, K., AND WILLIAMSON, J. R. Mathematical analysis of isotope labeling in the citric acid cycle with applications to ^{13}C NMR studies in perfused rat hearts. *J. Biol. Chem.* **258**, 13785–13794 (1983).
10. BARTHOLOMAY, A. F. Stochastic models for chemical reactions. I. Theory of the unimolecular reaction process. *Bull. Math. Biophys.* **20**, 175–190 (1958).
11. BARTHOLOMAY, A. F. A stochastic approach to statistical kinetics with application to enzyme kinetics. *Biochemistry* **1**, 223–230 (1962).
12. BARTHOLOMAY, A. F. Enzymatic reaction-rate theory: A stochastic approach. *Ann. NY Acad. Sci.* **96**, 897–912 (1962).
13. GILLESPIE, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
14. HIRSCH, M. W., AND SMALE, S. "Differential Equations, Dynamical Systems, and Linear Algebra." Academic Press, New York, 1974.
15. MOORE, W. J. "Physical Chemistry," 4th ed. Prentice–Hall, Englewood Cliffs, NJ, 1972.
16. STEINFELD, J. I., FRANCISCO, J. S., AND HASE, W. L. "Chemical Kinetics and Dynamics." Prentice–Hall, Englewood Cliffs, NJ, 1989.
17. FELLER, W. "An Introduction to Probability Theory and its Applications," Vol. 1, 3rd ed. Wiley, New York, 1968.
18. SOLOMON, F. "Probability and Stochastic Processes." Prentice–Hall, Englewood Cliffs, NJ, 1987.
19. ROSS, S. M. "Introduction to Probability Models," 4th ed. Academic Press, Boston, 1989.
20. STANDISH, T. A. "Data Structure Techniques." Addison–Wesley, Reading, MA, 1980.
21. MALLOY, C. R., SHERRY, A. D., AND JEFFREY, F. M. H. Evaluation of carbon flux and substrate selection through alternate pathways involving the citric acid cycle of the heart by ^{13}C NMR spectroscopy. *J. Biol. Chem.* **263**, 6964–6971 (1988).
22. OVADI, J., AND SRERE, P. A. Channel your energies. *Trends Biochem. Sci.* **17**, 445–447 (1992).
23. O'LEARY, M. H. Multiple isotope effects on enzyme-catalyzed reactions. *Annu. Rev. Biochem.* **58**, 377–401 (1989).

24. OLEFSKY, J. M., AND MOLINA, J. M. Insulin resistance in man. In "Ellenberg and Rifkin's Diabetes Mellitus, Theory and Practice," (H. Rifkin and D. Porte, Jr., Eds.), 4th ed., Chap. 8, Elsevier, New York, 1990.
25. MOREADITH, R. W., AND LEHNINGER, A. L. The pathways of glutamate and glutamine oxidation by tumor cell mitochondria. *J. Biol. Chem.* **259**, 6215-6221 (1984).
26. GOPHER, A., VAISMAN, N., MANDEL, H., AND LAPIDOT, A. Determination of fructose metabolic pathways in normal and fructose-intolerant children: A ^{13}C NMR study using $[\text{U-}^{13}\text{C}]$ fructose. *Proc. Natl. Acad. Sci. USA* **87**, 5449-5453 (1990).
27. DAVIS, R., AND KING, J. J. The origin of rule-based systems in AI. In "Rule-Based Expert Systems" (B. G. Buchanan and E. H. Shortliffe, Eds.), pp. 20-52. Addison-Wesley, Reading, MA, 1984.
28. PARK, S. K., AND MILLER, K. W. Random number generators: Good ones are hard to find. *Commun. ACM* **31**, 1192-1201 (1988).
29. MARSAGLIA, G. Remarks on choosing and implementing random number generators. *Commun. ACM* **36**, 105-108 (1993). (Technical correspondence).
30. SULLIVAN, S. J. Another test for randomness. *Commun. ACM* **36**, 108 (1993). (Technical correspondence)
31. PARK, S. K., MILLER, K. W., AND STOCKMEYER, P. K. Response. *Commun. ACM* **36**, 108-110 (1993). (Technical correspondence)