

Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets

PETER J. E. GOSS^{†‡} AND JEAN PECCOUD^{§¶||}

[†]Department of Organismic and Evolutionary Biology, Harvard University, Museum of Comparative Zoology Laboratories, Third Floor, 26 Oxford Street, Cambridge, MA 02138; and [§]Techniques de l'imagerie, de la Modélisation et de la Cognition/Institut d'Informatique et de la Mathématiques Appliquées de Grenoble, Institut Albert Bonnitot, Fawlté de Médecine de Grenoble, 38706 La Tranche cedex, France

Communicated by Herman Chernoff, Harvard University, Cambridge, MA, January 20, 1998 (received for review November 20, 1997)

ABSTRACT An integrated understanding of molecular and developmental biology must consider the large number of molecular species involved and the low concentrations of many species *in vivo*. Quantitative stochastic models of molecular interaction networks can be expressed as stochastic Petri nets (SPNs), a mathematical formalism developed in computer science. Existing software can be used to define molecular interaction networks as SPNs and solve such models for the probability distributions of molecular species. This approach allows biologists to focus on the content of models and their interpretation, rather than their implementation. The standardized format of SPNs also facilitates the replication, extension, and transfer of models between researchers. A simple chemical system is presented to demonstrate the link between stochastic models of molecular interactions and SPNs. The approach is illustrated with examples of models of genetic and biochemical phenomena where the ULTRASAN package is used to present results from numerical analysis and the outcome of simulations.

Many processes in molecular biology involve small numbers of molecules. Recent observations of gene expression in individual cells illustrate the stochastic nature of transcription (1–5). Multimodal probability distributions arise in a model of a single DNA molecule amplified by using PCR at a high amplification rate (6). Models of genetic networks are becoming increasingly important; for example, the lysis/lysogeny decision of lambda phage has been modeled both deterministically (7) and stochastically (8). A major difference between the deterministic and stochastic models is that the initial condition fixes the outcome in a deterministic model, but in a stochastic model, qualitatively distinct outcomes, such as lysis or lysogeny, can arise from identical initial conditions caused by the random timing of events (8, 9). Developmental decisions often may depend on small numbers of many different types of molecules, and stochastic effects cannot be ignored. A systematic effort to reduce variation in inbred mice reveals that most of the residual variation is attributed to events that occur at or before fertilization (10).

A theoretical understanding of molecular and developmental biology must account for the variation caused by stochastic interactions of molecules present in small absolute numbers and the large number of different types of molecules and complex feedback loops involved (11). Traditional mathematical tools are not well-suited to modeling the dynamic behavior of distributions of tens or hundreds of different types of interacting molecules. Such systems can, in principle, be represented by a series of coupled Kolmogorov equations (12). However, such equations are impossible to treat analytically

for all but the simplest cases of molecular interactions. Quantitative stochastic models can be used to integrate detailed biochemical data and to help understand the behavior of complex systems of molecular interactions.

We introduce an approach to modeling stochastic systems in molecular biology, using stochastic Petri nets (SPNs) (no relation to Petri dishes). SPNs are a formalism developed in the field of computer science and have a standard graphical representation, which is easy to interpret and to use for defining models (13). The graphical representation of molecular interactions as SPNs is similar to standard representations in biochemistry. Computer packages are available that integrate the graphical definition of SPNs with numerical analysis or simulation.** Models of stochastic molecular interactions can be represented as SPNs. The stochastic process that results from this representation is equivalent to that used for modeling stochastic chemical reactions (12), often referred to as the chemical master equation (15). Thus, it is possible to define complex models and quantitatively solve them for the probability distributions of molecular species by using SPN software without writing specific computer code for each model. The software makes it easy to modify or replicate models or to transfer models in a standardized format.

Even though the use of SPN software makes the implementation of a model much easier than the use of a low-level programming language, it is still far from being trivial. To understand the functionality of the software, a general background in stochastic process theory, statistics, and Monte Carlo simulation techniques is mandatory. Some understanding of the numerical algorithms used to solve the models is highly desirable to figure out the limitations of the software and to operate it effectively. Textbooks in this field are rare, but it might be useful to read ref. 16. The contents of this book should be accessible to anyone who plans to make use of the techniques described in this paper. An introduction to the theory of stochastic processes as applied to chemical and biological systems is given by refs. 17 and 18.

The paper is organized into three sections. The first section introduces the terminology of SPNs and defines the interpretation of SPNs we use to represent molecular interactions. The graphical representation of SPNs is illustrated by using a simple chemical system, model I. The second section describes the different analysis techniques available by using SPNs. As an example, the SPN representation of a simple genetic system, model II, is explored by using numerical analysis and simulation. The SPN approach is validated by comparing the results

Abbreviation: SPN, stochastic Petri net.

[‡]e-mail: goss@mcz.harvard.edu.

[¶]To whom reprint requests should be addressed. e-mail: jean.peccoud@imag.fr.

^{||}Present address: Department of Electrical Engineering, Box 352500, University of Washington, Seattle, WA 98195-2500.

**A general description of SPNs and references for further information can be found at <http://www.dsi.unimi.it/Users/Tesi/trompede/petri/home.html>. Details of ULTRASAN can be found at <http://chaos.crhc.uiuc.edu/PERFORM>, and from ref. 14.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/956750-6\$2.00/0
PNAS is available online at <http://www.pnas.org>.

of this model to results derived symbolically for the same system (19). The third section presents a more realistic biological example. Model III is a stochastic model of ColE1 plasmid replication, derived from a deterministic model by Brendel and Perelson (20). Simulation results are presented for model III and compared with results from the deterministic model. The discussion outlines the major advantages of using SPNs for stochastic models of molecular interactions as well as various issues that need to be addressed.

Representation of Molecular Interactions as SPNs

A number of variations on the definition of a SPN and its extensions are present in the literature. We give an abbreviated introduction to SPNs and show how to represent molecular interactions as SPNs via a strict interpretation of a limited class of SPNs and an additional constraint based on the kinetics of chemical reactions. First, we define the components of an SPN, following the terminology of ref. 13, and the molecular interaction interpretation of this terminology (Table 1):

An SPN comprises a set of places P , a set of transitions T , an input function I , an output function O , a weight function W , and an initial marking M_0 .

To represent a system of molecular interactions as an SPN, each place represents a distinct molecular species. Places contain tokens, which represent individual molecules. The number of tokens in a place p_i is its marking. The state of the system is given by a vector M , the global marking. The initial marking M_0 is the number of molecules of each species in the system at time $t = 0$.

Each transition represents an elementary chemical reaction. Input and output functions link places and transitions, and determine the stoichiometric coefficients of the molecular species involved in the reaction. The rate of the reaction is represented by the weight function. In the graphical representation of SPNs, places are drawn as circles, and transitions as rectangles (13). Input and output functions are drawn as arrows, referred to as directed arcs, linking input and output places to transitions, where input places represent reactants and output places products. Coefficients in the reaction equation greater than one are drawn as arcs labeled with this coefficient. Model I is an SPN representation of the dimerization reaction $2R \rightleftharpoons R_2$ (Fig. 1). Note the similarity of this figure with those found in textbooks in biochemistry. Also, this is the representation of the reaction by its Vol'pert graph, used in chemical kinetics.

A transition is said to be enabled when the markings of all of its input places are at least as great as the coefficients of their respective input arcs. For example, the transition t_+ in Fig. 1 is enabled if and only if $m_{\text{monomer}} \geq 2$. In our interpretation of SPNs as models of molecular interactions, this is equivalent to the statement that a reaction can occur only if sufficient reactant molecules are present. Enabled transitions can fire, representing a single molecular reaction event where reactant molecules are removed and product molecules added according to the coefficients of the reaction. Thus, when transition t_+

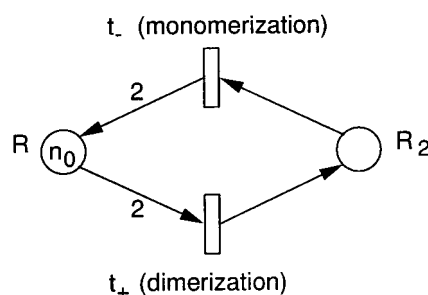


FIG. 1. SPN representation of the dimerization reaction $2R \rightleftharpoons R_2$. The transition t_+ represents dimerization, with weight function $w_{t_+} = c_+ m_{\text{monomer}}(m_{\text{monomer}} - 1)$, where c_+ is a constant. The transition t_- represents dissociation, with weight function $w_{t_-} = c_- m_{\text{dimer}}$. When no initial marking is specified for a place, it is assumed to be zero. Thus, the initial marking is $M_0 = \{n_m, 0\}$.

fires, m_{monomer} is decreased by two and m_{dimer} increased by one, representing a single dimerization event. Now we must show that the stochastic process by which transitions fire is equivalent to that for chemical reactions.

In SPNs, enabled transitions fire with an exponentially distributed time delay. The rate parameter for each transition is given by the weight function W , and may, in general, be a function of the global marking M (13). Thus, SPNs are a class of Markov jump processes with discrete state space. Chemical reactions at low concentration also are modeled by Markov jump processes (15) where the rate of each chemical reaction usually is constructed from the stoichiometry of the reaction. The stochastic reaction rate of a chemical reaction is a function of only those molecular species involved as reactants or catalysts, and a stochastic rate constant c , which takes into account volume, temperature, pH, and other environmental factors. The stochastic rate constant is related to the deterministic kinetic constant k (15). In the limit as the number of molecules tends to infinity, the stochastic rate of a reaction is equal to the deterministic rate (12). For SPNs to represent molecular interactions exactly, the rate constant w_j for a transition t_j must be the stochastic rate of the reaction the transition represents.

For monomolecular reactions, the stochastic and deterministic rate constants are equal. Thus the stochastic rate of the reaction $R_2 \rightarrow 2R$, represented by transition t_- in Fig. 1, is $w_- = k_- m_{\text{dimer}}$, where m_{dimer} is the number of R_2 dimers in the system and k_- is the dissociation rate constant. When the order of the chemical reaction is greater than one, the relationship of the stochastic reaction rate to the deterministic reaction rate depends on the volume of the system and the numbers of each reactant required for the reaction (12, 15). The rate of transition t_+ in Fig. 1 is given by $w_{t_+} = c_+ m_{\text{monomer}}(m_{\text{monomer}} - 1)$, where m_{monomer} is the number of R monomers in the system, $c_+ = k_+ / V N_A$, k_+ is the deterministic dimerization rate constant, V is the volume of the reaction system and N_A is Avogadro's number.

The representation of molecular interactions as SPNs provides a well-defined formalism. Any chemically reacting system can be represented as an SPN in this way. The stochastic process arising from the SPN representation is equivalent to that arising from the chemical equations written to describe the system. Put in another way; if we were to write down the Kolmogorov equations for the SPN, the result would be the chemical master equation for the system of molecular interactions represented.

There are a number of extensions to SPNs that will simplify modeling of complex systems of molecular interactions. For example, generalized SPNs may have instantaneous transitions, which fire as soon as they are enabled. Some extensions of SPNs allow the possibility of generalized distributions of time delays, including delays of fixed length called determin-

Table 1. Molecular interpretation of SPN terminology

SPN term	Molecular interpretation
Place	Molecular species
Token	Molecule
Marking	Number of molecules
Transition	Reaction
Input place	Reactant
Output place	Product
Weight function	Rate of reaction
To be enabled	For a reaction to be possible
To fire	For a reaction to occur

istic transitions (13). It is also possible to put additional restrictions on transitions. For example, the software we use in this paper, ULTRASAN, uses an extension of SPNs called stochastic activity networks (SANs). SANs include input gates to control the conditions under which transitions are enabled and output gates to control the effect of transitions (21).

Analysis of SPN Models of Molecular Interactions

Three broad approaches to the analysis of SPNs may be differentiated. The first approach uses structural analysis of the Petri net underlying the SPN (13), where the transitions of the SPN are converted into instantaneous transitions to form a Petri net. This approach to analysis is similar to the problem of classifying states in a Markov chain into transient, fixed, or recurring states. Some of the possible uses of structural analysis of Petri net models of molecular interactions have been described (22, 23).

Second, numerical analysis can be used to derive both steady-state and transient behavior. Numerical analysis algorithms explicitly generate the Markov chain associated with the SPN (13), but require that the state space of the Markov chain be less than several hundred thousand states (24, 25). For many models of molecular interactions, however, the size of the state space increases very rapidly with the number of different types of molecules, and numerical analysis is impractical.

Third, algorithms are available for simulating both steady-state and transient behavior and estimating the distributions of results (21). Relative confidence intervals can be estimated and used interactively to determine how many runs of the simulation are required to produce a given level of precision.

Following Sanders (26), results may be associated with places, called reward measures, or with transitions, called impulse measures. In molecular terms, reward measures might be used to estimate the distribution of the number of molecules of some species at a particular time, or of the average number of molecules over some period of time. Impulse measures might be used to determine the number of times a reaction occurs in a particular time interval.

As an example of numerical analysis and simulation of an SPN, model II is a simple model of protein synthesis. Peccoud and Ycart (19) symbolically solved the Kolmogorov equations arising from this system, and we compare the results of numerical analysis and simulation using ULTRASAN (21) to those results.

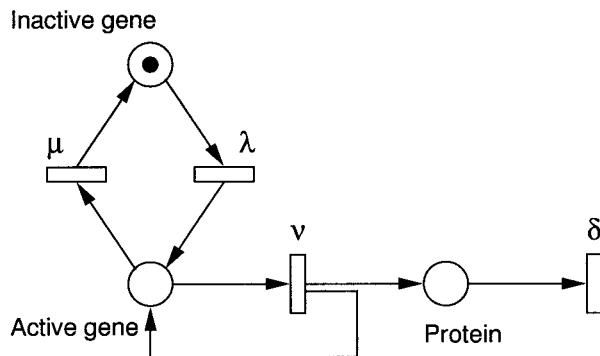


FIG. 2. SPN representation of a simple model of protein synthesis. The SPN contains three places: p_1 = inactive gene, p_2 = active gene, and p_3 = protein. The four transitions and their respective rate parameters are activation (λ), inactivation (μ), synthesis (ν), and degradation (δ). The dot inside place inactive gene represents a single token (a single copy of the gene). Thus, the initial marking of the SPN is $M_0 = \{1, 0, 0\}$. Other molecules in this system, such as RNA polymerase, are assumed to be in constant concentration and are not explicitly represented.

Model II contains a single copy of a gene, which is initially inactive, but which subsequently may be active or inactive. Protein may be produced when the gene is active and may be degraded at any time. The SPN representation of this model is given in Fig. 2. The place active gene is simultaneously an input and an output of the transition synthesis (ν), because it is required for synthesis but not consumed, acting effectively as a catalyst. Protein degradation is drawn as a transition with no output arc, representing a reaction with no products.

We define two output measures for this model. The first one is the probability distribution of the number of protein molecules. To render the state space of the SPN finite for numerical analysis, it is necessary to limit the number of protein molecules allowed in the system, by setting a reflecting boundary on the state space; if this boundary is set sufficiently high, the error in the approximation is negligible. In ULTRASAN, this constraint can easily be implemented by using an input gate. A second output measure estimates the probability that the number of proteins is at this limit. With the parameter values used in Table 2, this probability is of the order of 10^{-50} .

Table 2 presents results on the mean and variance of the number of proteins from both numerical analysis and simulation using ULTRASAN, for both transient and steady-state behavior. The results are consistent with the symbolic solution of Peccoud and Ycart (19), which validates the SPN approach. Peccoud and Ycart did not derive the analytic distribution of protein number for this model. Fig. 3 shows the distribution of protein number in steady state by using numerical analysis in ULTRASAN.

Stochastic Model of ColE1 Plasmid Replication

The biomolecular mechanisms underlying plasmid copy number control have been studied extensively (27). The spontaneous loss of plasmids is experimentally important, because bacteria without plasmids often replicate more quickly and can outcompete bacteria with plasmids (27). Therefore, it is important to be able to estimate not only the mean number of plasmids per bacterium, but also the variance in plasmid number and the probability of spontaneous loss of plasmids from a bacterial lineage.

Model III is an SPN model of plasmid ColE1 replication. It is the stochastic equivalent of a differential equation model based on deterministic kinetics (20). Fig. 4 shows the reaction

Table 2. Mean and variance of protein number from model II

	Protein number	
	Mean, \pm SE*	Variance, \pm SE*
Transient analysis (time $t = 10$)		
Symbolic solution	1.488	1.858
Numerical analysis	1.488	1.858
Simulation [†]	1.481 ± 0.004	1.852 ± 0.011
Transient analysis (time $t = 100$)		
Symbolic solution	7.202	8.334
Numerical analysis	7.202	8.334
Simulation [†]	7.171 ± 0.009	8.315 ± 0.039
Steady-state analysis		
Symbolic solution	8.333	9.487
Numerical analysis	8.333	9.487
Simulation [‡]	8.333 ± 0.031	9.551 ± 0.100

Results of symbolic solution are from ref. 19. Numerical analysis and simulation results are from an SPN model defined and solved by using ULTRASAN. Parameter values: activation rate, $\lambda = 1$; inactivation rate, $\mu = 5$; synthesis rate, $\nu = 1$; and degradation rate, $\delta = 0.02$. Maximum number of protein molecules = 100.

*Standard error of variables estimated using simulation.

[†]Based on 10^5 runs (computation time on a Pentium 166 = 386 sec).

[‡]Based on 10^3 runs (computation time on a Pentium 166 = 686 sec).

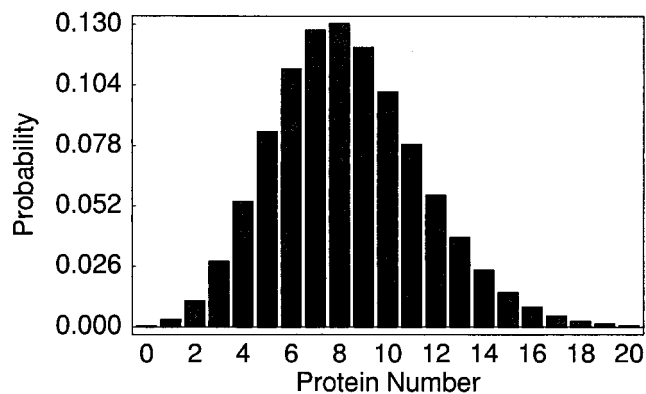


FIG. 3. Distribution of protein number in simple model of gene product synthesis. The distribution was generated from steady-state numerical analysis using the same parameter values as Table 2.

network of ColE1 replication, modified from fig. 1 of Brendel and Perelson (20) with permission. Full details of the ULTRASAN implementation of this model are available on request.

The ColE1 replication system contains 18 molecular reactions between 10 different molecular species, including seven different plasmid complexes and RNA I, RNA II, and Rom protein. The deterministic model representing this system consists of 10 coupled differential equations with 19 rate constants (20). The concentrations of plasmid, RNA I, RNA II, and Rom protein were derived numerically for steady state (20).

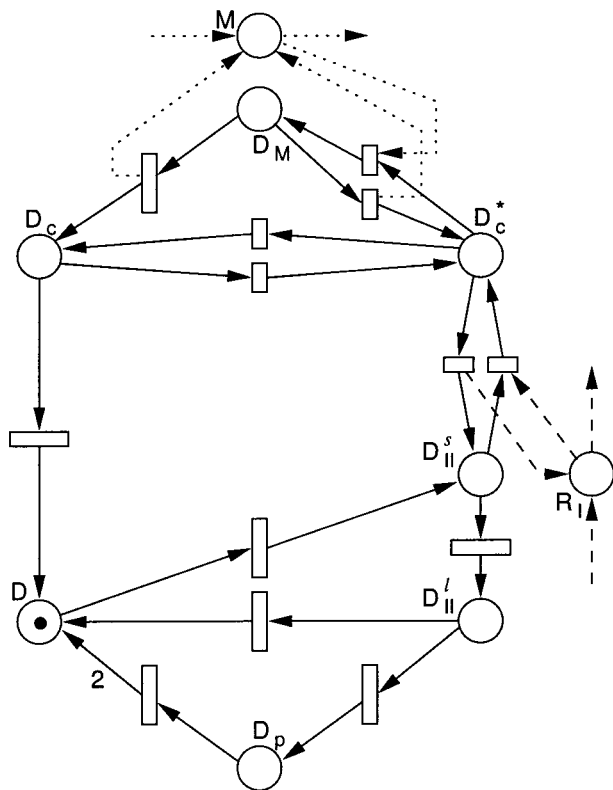


FIG. 4. SPN representation of plasmid ColE1 replication system. The reactions represented and the notation used are from fig. 1 of Brendel and Perelson (20). Plasmid DNA occurs in free form (D), or in complexes with RNA II (D_{II}^s , D_{II}^i , and D_p), with RNA II and RNA I (D_c^* and D_c), or with RNA II, RNA I, and Rom protein (D_M). Replication occurs when primed plasmid DNA (D_p) is converted to free DNA (D). Free RNA I and Rom protein are represented by places R_I and M , respectively. Initial marking of free plasmid is 1. (Modified with permission from ref. 20.)

Using ULTRASAN we model the number of each molecular species in each bacterial lineage. During each generation the volume of a bacterium grows exponentially from V_0 to $2V_0$, in doubling time $\tau_D = 80$ min. Continuous volume growth is approximated by dividing the bacterial generation into small deterministic time steps; increasing the number of time steps does not affect the results of the model (data not shown). Most simulations start with a single plasmid, to simulate the behavior of a newly invading plasmid.

Fig. 5 shows the changing distribution of plasmid copy number at 10-min intervals through the first generation. Not only does the mean number of plasmids per bacterium increase, but also the variance. At the end of the first generation, mean plasmid number is 18.6 and the SD of plasmid number is 5.4. Of biological interest is the probability of no replication during the first generation, or, equivalently, the frequency of exactly one plasmid being present immediately before bacterial division, estimated at $4.0 \cdot 10^{-4}$ [95% confidence interval ($1.3 \cdot 10^{-4} - 1.1 \cdot 10^{-3}$)]. If there is only one copy of the plasmid at the end of the first generation, one of the daughter cells cannot inherit a plasmid. Thus, this probability is related to the probability of the plasmid establishing itself in a bacterial population.

We also simulated the change in plasmid copy number over several generations. At the end of each generation, the number of molecules of each species was divided in two to form the next generation. Thus, each generation begins with a distribution of plasmids based on the distribution from the previous generation. Segregation of plasmids during bacterial replication often may be random in practice (27–30).

Fig. 6 shows plasmid number per bacterium for the first 10 generations, starting with one or 100 plasmids. Because the volume of bacteria changes with time, the number of plasmids per bacterium is cyclic. Plasmid number was sampled every 20 min during the 80-min bacterial doubling time. Table 3 shows the number of plasmids, as well as the number of molecules of RNA I and Rom, at the midpoint of generations 8, 9, and 10 starting from a single plasmid. The means for each species from the stochastic model are very similar to the steady-state values of the deterministic model.

After 6–8 generations, plasmid number appears to approach an asymptotic periodic distribution, irrespective of whether the simulations were initiated with one or 100 plasmids per bacterium. The distributions of plasmid number starting with one or 100 plasmids are indistinguishable by the end of generation 8 ($\chi^2 = 34.3$, 25 degrees of freedom, $P > 0.1$),

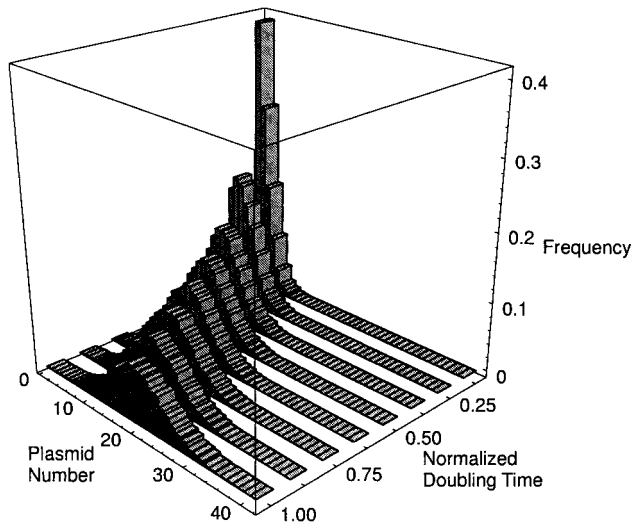


FIG. 5. Distribution of plasmid copy number in a single bacterium during first generation with 80-min doubling time.

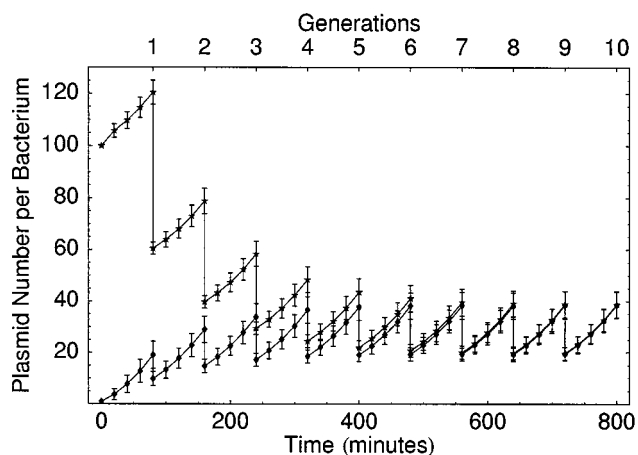


FIG. 6. Plasmid copy number during 10 generations starting from 1 or 100 plasmids with 80-min doubling time. Mean and SD of plasmid number is measured every 20 min and at the beginning and end of each generation.

indicating that the asymptotic distribution has been approached (data not shown).

Discussion

The stochastic representation of molecular interactions has a stronger theoretical basis than deterministic kinetics when the number of molecules is small (15, 31). Both the quantitative and the qualitative results of a stochastic model may differ from the deterministic equivalent (15). Stochastic effects may be crucial when dealing with gene expression, and the stochastic nature of transcription has been experimentally observed in a number of cases (3–5). Stochastic effects also may be important at the level of signal transduction (32–34), cell replication (35), or cellular differentiation (1, 2). Quantitative information is essential for understanding these phenomena (36, 37). Detailed stochastic models may be used to integrate information from molecular biology and genetics, to test hypotheses about underlying mechanisms, to guide experimentation, or to estimate the rates of processes that are difficult to observe directly.

The importance of the underlying stochastic nature of molecular interactions has been recognized in chemistry for many years (12, 38). There is a large body of literature describing how to derive stochastic processes for chemical reactions at low concentration (15, 17, 18), as well as for simple biochemical processes (39–45). The Kolmogorov equation resulting from these stochastic processes, often called the chemical master equation, is analytically intractable for all but the simplest cases (15, 18). Most of the complex interactions encountered in biological systems fall well beyond the scope of exact analysis. An exact algorithm is available for simulation of

coupled chemical reactions (46). However, efficient implementation of this algorithm for complex biological models requires considerable skill, as does the estimation of confidence intervals for the mean and the variance of desired quantities. Programs written to perform such simulations efficiently often are limited to specific models and can be hard to understand and thus difficult to replicate or to modify.

Nevertheless, there have been a number of papers presenting models of stochastic effects caused by low molecular concentration at the level of the metabolic effect of dioxygen free radicals (47), the lysis/lysogeny decision of phage lambda (8), signal transduction (32–34), gene induction (9, 19, 48, 49), bacterial chemotaxis (50–52), and cellular selection (53–57). Stochastic modeling of complex molecular interactions is being driven by advances in molecular biology and facilitated by the increasing availability of powerful computers to analyze such models. However, research in this field is being greatly hampered by the lack of a consistent formalism and the difficulties of case-by-case simulation.

Ideally, when a new modeling approach is introduced, it should contain a formalism that justifies the theoretical basis for the approach, and facilitates the representation of models. Petri nets, and their stochastic extensions, are well-defined formalisms. The graphical representation of stochastic molecular interactions as Petri nets or SPNs is readily applicable to complex metabolic and genetic networks. The ability of Petri nets to represent complex systems of interactions has suggested their use for qualitative analysis of metabolic processes and networks (22, 23). However, without their stochastic extensions, Petri nets cannot represent time as a continuous variable, and, therefore, are unable to model chemical kinetics (23). Qualitative analysis of biochemical and genetic networks may still be useful, for example, for identifying steady states in the absence of reliable kinetic data (58), for lumping of large systems (22), or for the identification of metabolic bottlenecks (23).

The representation of molecular interactions as SPNs we have described is justified by the equivalence of the underlying stochastic process to that of the Kolmogorov equations of the chemical system. The formalism of SPNs is well defined and the use of SPNs is validated by an extensive body of work in computer science. The graphical representation of molecular interactions as SPNs is similar to representations commonly used in molecular biology. Complex systems are much easier to represent graphically using SPNs than with the chemical master equation. Where numerical analysis is not feasible, simulation algorithms have been designed and optimized. Computer packages are available that integrate the graphical representation of SPN models with their analysis, giving biologists ready access to a wide range of tools for modeling molecular interactions as SPNs. The package we use, ULTRASAN, is designed to cope with numerical analysis or simulation of stiff systems, where the relative range of time constants is large (14). Further details of the techniques used by UltraSAN for numerical analysis and simulation are provided with the software (14).

An important benefit of the formalism of SPNs is that models can more easily be replicated or extended than if they were implemented by using case-by-case simulation. All of the SPN models described in this paper are available (<http://www-timc.imag.fr/spns>) and may be used by anyone who has access to the ULTRASAN package. If realistic models of networks of molecular reactions are to become widely used in molecular biology, a systematic framework within which to build and extend models is vital.

Two important questions need to be addressed in any stochastic model of molecular interactions. How should we model the interaction between deterministic kinetics, associated with continuous variables such as the concentration of small molecules, and stochastic kinetics, associated with dis-

Table 3. Asymptotic plasmid, RNA I, and Rom copy number

	Plasmid		RNA I		Rom	
	Mean	SD	Mean	SD	Mean	SD
Deterministic model	28	N/A	450	N/A	750	N/A
Stochastic model*						
Generation 8	27.7	3.4	446	55	745	92
Generation 9	28.0	3.4	450	55	753	91
Generation 10	27.8	3.4	448	54	748	91

Results of deterministic model are from ref. 20, table 1, WT Rom⁺. Results of stochastic SPN model are measured at the midpoint of generations 8, 9, and 10. N/A, not applicable.

*Based on 10³ runs of transient simulation starting with one plasmid.

crete variables such as plasmid copy number? Theoretically, this type of interaction may be represented as a Markov jump process with drift (18). However, more work is required to learn how to model such processes in practice.

Second, how do we cope with unknown parameter values? Kinetic models, whether stochastic or deterministic, require kinetic parameters, but these are often difficult to measure. Detailed and realistic stochastic models may be able to help. Molecular biology is excellent at observing qualitative behavior. The resultant behavior of models provides constraints on the range of parameter values. Optimization techniques for estimating such parameter values in conjunction with SPNs would be extremely useful in this regard, although this is a difficult problem.

One limitation of using SPNs for modeling biochemical systems is the lack of a spatial dimension. Compartmentalized systems present no difficulty, because they can be represented as separate, but interacting, systems, with the movement of molecules between compartments represented by transitions. However, in the current definition of SPNs it is difficult to see how a true reaction-diffusion model could be implemented.

In this paper we have considered models of stochastic molecular interactions. However, random phenomena at many levels of biology share similarities with molecular interactions. With an interpretation of the basic components of SPNs appropriate for these different levels, SPNs have potential applications for stochastic modeling in many areas of biology, including embryology, developmental biology, population biology, and ecology.

For the purpose of illustration, modeling in biology may be divided into four overlapping processes: model design, representation, analysis, and interpretation. Detailed biological understanding, along with an understanding of the underlying stochastic process, is required to design a realistic model and interpret it appropriately. The approach presented in this paper deals with the representation and analysis of stochastic models as SPNs. By using this approach, complex molecular or genetic networks may be modeled using currently available software packages, allowing biologists to concentrate on the design and interpretation of detailed models rather than their implementation.

We thank B. Ycart for introducing us to the notion of SPNs. We thank R. C. Lewontin and J. Demongeot for their support. We thank A. Barbadilla, H. Chernoff, A. Conway, W. Gilbert, R. Helme, C. Jacob, U.-L. Pen, M. Siegal, J. Tóth, D. Weinreich, and anonymous reviewers for careful readings of the manuscript and helpful suggestions. This work was supported by the Institut National de la Santé et de la Recherche Médicale (J.P.) and a travel grant from the Department of Organismic and Evolutionary Biology, Harvard University (P.J.E.G.).

- Baroffio, A. & Blot, M. (1992) *J. Cell Sci.* **103**, 581–587.
- Abkowitz, J. L., Catlin, S. N. & Guttorp, P. (1996) *Nat. Med.* **2**, 190–197.
- Weintraub, H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5819–5823.
- Ko, M. S. H., Nakauchi, H. & Takahashi, N. (1990) *EMBO J.* **9**, 2835–2842.
- Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M. & Martin, D. I. K. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7125–7129.
- Peccoud, J. & Jacob, C. (1996) *Biophys. J.* **71**, 101–108.
- McAdams, H. H. & Shapiro, L. (1995) *Science* **269**, 650–656.
- Arkin, A., Ross, J. & McAdams, H. H. (1998) *J. Mol. Biol.*, in press.
- McAdams, H. H. & Arkin, A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 814–819.
- Gärtner, K. (1990) *Lab. Anim.* **24**, 71–77.
- Fagerström, T., Jagers, P., Schuster, P. & Szathmari, E. (1997) *Science* **274**, 2039–2040.
- Kurtz, T. G. (1972) *J. Chem. Phys.* **57**, 2976–2978.
- Ajmone Marsan, M., Balbo, G., Chiola, G., Conte, G., Donatelli, S. & Franceschinis, G. (1991) *Microelectron. Reliab.* **31**, 699–725.
- Center for Reliable and High-Performance Computing *ULTRASAN User's Manual* (Univer. of Illinois at Urbana-Champaign), Version 3.0.
- Érdi, P. & Tóth, J. (1989) *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models* (Princeton Univ. Press, Princeton).
- Stewart, W. J. (1994) *Introduction to the Numerical Solution of Markov Chains* (Princeton Univ. Press, Princeton).
- Van Kampen, N. G. (1981) *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).
- Gardiner, C. W. (1990) *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences* (Springer, New York), 2nd Ed.
- Peccoud, J. & Ycart, B. (1995) *Theor. Popul. Biol.* **48**, 222–234.
- Brendel, V. & Perelson, A. S. (1993) *J. Mol. Biol.* **229**, 860–872.
- Sanders, W. H., Obal II, W. D., Qureshi, M. A. & Widjanarko, F. K. (1995) *Perform. Eval.* **24**, 89–115.
- Reddy, V. N., Liebman, M. N. & Mavrouniotis, M. L. (1996) *Comput. Biol. Med.* **26**, 9–24.
- Hofestädt, R. (1994) *J. Syst. Anal. Model. Simul.* **16**, 113–122.
- Deavours, D. D., Obal II, W. D., Qureshi, M. A., Sanders, W. H. & van Moorsel, A. P. A. (1995) in *Proceedings of the Sixth International Workshop on Petri Nets and Performance Models, Durham, NC* (IEEE Comput. Soc. Press, Los Alamitos, CA), pp. 216–217.
- Sanders, W. H. & Meyer, J. F. (1991) *IEEE J. Select. Areas Commun.* **9**, 25–36.
- Sanders, W. H. & Meyer, J. F. (1991) in *Dependable Computing for Critical Applications*, eds. Avizienis, A. & Laprie, J. (Springer, New York), Vol. 4, pp. 215–237.
- Summers, D. K. (1996) *The Biology of Plasmids* (Blackwell, Oxford).
- Womble, D. D. & Rownd, R. H. (1986) *J. Mol. Biol.* **191**, 367–382.
- Womble, D. D. & Rownd, R. H. (1986) *J. Mol. Biol.* **192**, 529–548.
- Rosenfeld, R. & Grover, N. B. (1993) *Plasmid* **29**, 94–116.
- Gillespie, D. T. (1992) *Physica A* **188**, 404–425.
- Lamb, T. D. (1994) *Biophys. J.* **67**, 1439–1454.
- Lamb, T. D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 566–570.
- Felber, S., Breuer, H. P., Petruccione, F., Honerkamp, J. & Hofmann, K. P. (1996) *Biophys. J.* **71**, 3051–3063.
- Michaelson, J. (1993) *Lab. Invest.* **69**, 136–151.
- Maddox, J. (1994) *Nature (London)* **368**, 95.
- Loomis, W. F. & Sternberg, P. W. (1995) *Science* **269**, 649.
- Delbrück, M. (1940) *J. Chem. Phys.* **8**, 120–124.
- Bartholomay, A. F. (1959) *Ann. N.Y. Acad. Sci.* **96**, 897–912.
- Smeach, S. C. & Gold, H. J. (1975) *J. Theor. Biol.* **51**, 59–78.
- Smeach, S. C. & Gold, H. J. (1975) *J. Theor. Biol.* **51**, 79–96.
- Arányi, P. & Tóth, J. (1977) *Acta Biochim. Biophys. Sci. Hung.* **12**, 375–388.
- Hasstedt, S. (1978) *Biosystems* **10**, 319–328.
- Jacob, C., Breton, N., Daegelen, P. & Peccoud, J. (1997) *J. Chem. Phys.* **107**, 2913–2919.
- Jacob, C., Breton, N. & Daegelen, P. (1997) *J. Chem. Phys.* **107**, 2903–2912.
- Gillespie, D. T. (1977) *J. Phys. Chem.* **81**, 2340–2361.
- Moniz-Barreto, P. & Fell, D. A. (1993) *Biochem. Soc. Trans.* **21**, 256S.
- Ko, M. S. H. (1991) *J. Theor. Biol.* **153**, 181–194.
- Ko, M. S. H. (1992) *BioEssays* **14**, 341–346.
- Bray, D., Bourret, R. B. & Simon, M. I. (1993) *Mol. Biol. Cell* **4**, 469–482.
- Hauri, D. C. & Ross, J. (1995) *Biophys. J.* **68**, 708–722.
- Morton-Firth, C. J. & Bray, D. (1998) *J. Theor. Biol.*, in press.
- Till, J. E., McCulloch, E. A. & Siminovitch, L. (1964) *Proc. Natl. Acad. Sci. USA* **51**, 29–36.
- Korn, A. P., Henkelman, R. M., Ottensmeyer, F. P. & Till, J. E. (1973) *Exp. Hematol. (Charlottesville, VA)* **1**, 362–375.
- Nakahata, T., Gross, A. J. & Ogawa, M. (1982) *J. Cell. Physiol.* **113**, 455–458.
- Kobayashi, T. & Nakahata, T. (1989) *J. Cell. Physiol.* **138**, 24–28.
- Novak, J. P. (1993) *Blood Cells* **19**, 435–460.
- Thomas, R., Thieffry, D. & Kaufman, M. (1995) *Bull. Math. Biol.* **57**, 247–276.