Abstract Most biological systems are by nature hybrids consist of interacting discrete and continuous components, which may even operate on different time scales. Therefore, it is desirable to establish modeling frameworks that are capable of combining deterministic and stochastic, discrete and continuous, as well as multi-timescale features. In the context of molecular systems biology, an example for the need of such a combination is the investigation of integrated biological pathways that contain gene regulatory, metabolic and signaling components, which may operate on different time scales and involve on-off switches as well as stochastic effects. The implementation of integrated hybrid systems is not trivial because most software is limited to one or the other of the dichotomies above. In this study, we first review the motivation for hybrid modeling. Secondly, by using the example of a toggle switch model, we illustrate a recently developed modeling framework that is based on the combination of biochemical systems theory (BST) and hybrid functional Petri nets (HFPN). Finally, we discuss remaining challenges and future opportunities.

Keywords biochemical systems theory, Petri net, hybrid modeling, hybrid functional Petri net, toggle switch, canonical modeling, stochastic delay

1 Introduction

Several generic modeling frameworks have been developed during the past decades to predict the behavior of continuous, deterministic systems in biology and medicine. Among the best known is biochemical systems theory (BST) [1−5], which was designed for dynamic and steady-state systems, and metabolic control analysis [6−8], which was originally limited to responses at the steady-state but later augmented with the lin-log model formulation for dynamic analyses [9−11]. Most dynamical approaches implicitly assume that relatively large numbers of molecules interact freely within a well-mixed medium. This assumption is important because it permits the application of methods of statistical mechanics, which allows averaged, continuous rates, and thus the use of differential equations. However, many intracellular behaviors are discrete and some are apparently random in nature, for instance, because of low substrate concentrations or heterogeneous reaction environments. If these aspects cannot be validly ignored, the foundations of pure ordinary differential equation (ODE) representations present a significant limitation. As an alternative, some more recent modeling languages have been proposed to overcome this limitation. Examples include stochastic Petri nets [12,13] and stochastic automata [14]. The stochastic approaches pose their own challenges and are usually less efficient than differential equation methods [15,16]. The question thus arises of whether it is possible to combine the positive aspects of continuous, deterministic, and stochastic formulations and merge them into a unified, hybrid formalism. In this article, we discuss a promising combination of this type. We first identify more specifically the needs in integrative biological systems modeling, then outline a recently proposed, combined methodology, based on BST and hybrid functional Petri net (HFPN) [17], and finally point out future directions and opportunities. Additional aspects and technical details, along with different types of examples, are presented elsewhere [18].
2 Rationale for hybrid modeling

2.1 Stochasticity of biochemical reactions in vivo

Deterministic modeling relies fundamentally on the law of mass action. This law is derived from the assumption that thousands of molecules interact within a well-mixed and homogeneous medium, which in some sense admits the laws governing ideal gases. Supposing that molecular reactions are essentially random processes, one infers that macroscopic systems with a large number of interacting molecules allow the randomness to be averaged out so that the overall macroscopic state of the system can be accurately approximated by the deterministic laws. However, the population sizes of some components in an individual cell are limited (sometimes numbering only a few hundreds), and reactants often interact on a small scale and within non-homogeneous cellular environments. These observations render the assumption of continuous and deterministic approaches questionable. A typical example is the regulation of gene expression where only a few transcription factors interact with DNA binding sites in the gene’s regulatory sequence. Indeed, fluorescent probes have identified fluctuations at the level of individual cells [19,20], and one must wonder whether it is prudent and valid if these are averaged. Similarly, McAdams and Arkin [21] showed that low copy numbers of expressed RNAs can be significant for the regulation of downstream pathways. Moreover, cellular environments are highly compartmental and structured, which is far from the homogenous, well-mixed solutions that are typical of in vitro experiments [22]. A high degree of molecular crowding and the presence of endogenous obstacles in cellular media have important consequences for the thermodynamics within the cell [23,24] and strongly affect the diffusion processes [25]. For instance, the viscosity of the mitochondrion is 25–37 times higher than that of a typical in vitro experimental buffer [26]. Diffusion of macromolecules in the cytoplasm can be 5–20 times lower than in saline solutions [27]. Finally, rather than happening in a 3-dimensional volume, many reactions occur on two-dimensional membranes or in quasi-one-dimensional channels [28,29]. Thus, caution is needed when assumptions are made with regard to homogeneity and well mixing in biochemical reaction systems in vivo.

Stochastic approaches try to capture the inherently random nature of molecular collisions and to construct probabilistic models of the reaction kinetics [30,31]. If successful, such an approach is thus inherently more appropriate than averaging methods for the small, heterogeneous environments that are typical of in vivo conditions [32]. In spite of their advantages, stochastic models are still not sufficient for small-scale biological systems, because they do not explicitly account for spatial heterogeneity and are difficult to implement analytically. Furthermore, stochastic methods present major, genuine challenges. For instance, their construction requires detailed biochemical knowledge, including kinetic rates and numbers of molecules. In later sections we shall discuss to what degree HFPN is a suitable framework for implementing, simulating, and analyzing stochastic effects in biological pathway systems.

2.2 Accounting for multiple time scales in biological systems

Most models implicitly assume that the processes in biological systems run at similar time scales. If this is true, the dynamics of a system by and large depends on its current state and can be formulated with ordinary differential equations. In particular, the homogeneity of time scales implies that delay effects can be ignored. However, in reality, fast reactions (dimerization, phosphorylation, protein-DNA binding or unbinding) and slow reactions (transcription, translation, degradation) occur simultaneously, and they affect each other in multiple ways that critically influence their transient dynamics. For instance, it has been shown that delay within a feedback loop of mRNA transcription and protein expression can result in oscillatory expression patterns [33, 34]. Moreover, the combination of delays and positive feedback loops has been recognized as a mechanism that is able to maintain oscillatory expression patterns in noisy systems [35]. In the absence of a unifying modeling framework, any analysis of such a combination of stochastic and deterministic features has to be addressed through programming from scratch, which is seldom done.

2.3 Combined models of continuous and discrete components in intracellular systems

Biological systems often involve continuous and discrete phenomena side by side. For instance, gene transcription is switched on or off depending on the expression levels of other genes and on the presence or absence of transcription factors in sufficient quantities. The more or less continuous change in the concentration of a protein may trigger a discrete transition, such as the onset of mitosis or cell differentiation, which in turn changes the protein concentration. Hybrid methods easily address both discrete and continuous events within the same model. A hybrid approach is usually less computationally expensive than an exact discrete-event simulation and more accurate than continuous approximations, preserving the discrete or stochastic nature of the