Journal of Simulation (2007) 1, 81–96

www.palgrave-journals.com/jos

Discrete event modelling and simulation in systems biology

R Ewald[1]*, C Maus[1], A Rolfs[2] and A Uhrmacher[1]

[1]*Modeling and Simulation Laboratory, University of Rostock, Rostock, Germany*[2]*Neurobiological Laboratory, University of Rostock, Rostock, Germany*

With Systems Biology, a promising new application area for modelling and simulation emerges. Today's biologists are facing huge amounts of data delivered at different levels of detail by a multitude of advanced experimentation techniques. The Systems Biology approach copes with this information by cycling through phases of forming hypotheses, constructing models, experimenting with or analysing these models, and validating the findings by wet-lab experiments. A crucial point is therefore the way in which the knowledge about a system is formalized, that is, how a biological system is described, as this constrains the perception of the system as well as the scope of possible answers the model might provide. In this article, we compare different discrete event modelling formalisms (PETRI NETS, Stochastic π-CALCULUS, STATECHARTS, and DEVS) regarding their applicability to a cell biological system of current research interest, the Wnt signalling pathway. We then introduce the popular Gillespie algorithm, which is the foundation of many discrete event simulators for molecular-biological systems, and elaborate on some interesting extensions.
Journal of Simulation (2007) **1,** 81–96. doi:10.1057/palgrave.jos.4250018

Keywords: discrete event modelling formalisms; discrete event simulation; systems biology; stochastic simulation; PETRI NETS; Stochastic π-CALCULUS

1. Introduction

The goal of Systems Biology is to analyse the behaviour and interrelationships of functional biological systems (Kitano, 2002). Systems Biology is characterized by combining wet- and dry-lab experiments. As Systems Biology progresses, modelling and simulation is on its way to be established as one of the core tools in cell biological studies, which also inspires the development and use of different modelling and simulation methods.

Continuous systems models still prevail in Systems Biology (de Jong, 2002). This holds true for the sub-molecular scale, which relies on natural laws and is typically modelled based on partial differential equations, and it is also true for the level beyond molecules, for example, when focusing on concentrations of molecules. In the latter case, a continuous description by differential equations naturally supports a macro-view on the system of interest. The focus is on concentrations and their changes over time. Differential equations are particularly suitable for dynamics that occur continuously, evolve in a deterministic manner, progress at more or less the same speed, and can be easily described by real-valued variables. Extensions can relax these constraints, for example, by introducing delays, stochasticity or discreteness. All in all, continuous formalisms are suitable to describe the concentration dynamics of homogeneous cellular compartments that contain large numbers of involved entities.

Nevertheless, stochastic discrete event modelling and simulation is recently gaining ground as well. The discrete event view combines a continuous-time base with describing the dynamics of a system by distinguished state changes, that is, events that are triggered by the flow of time or the situation (Zeigler *et al*, 2000). Discrete event approaches address specific constraints of continuous and deterministic models: concentrations do not necessarily change continuously, particularly if the dynamics of a small amount of entities, like DNA molecules and plasmids, are modelled (Kuo and Keasling, 1996). In addition, the dynamics of some biological systems can be best approached in a stochastic manner, for example, if the gene regulation is to be described (Puchulka and Kierzek, 2004), where stochastic fluctuations are abundant (Cowan, 2003). Therefore, many state-of-the-art simulation systems in Systems Biology, for example (Van Gend and Kummer, 2001; Takahashi *et al*, 2002; Ramsey *et al*, 2005; http://www.mathworks.com/products/simbiology/, accessed 27 February 2007), offer to execute reaction networks by numerical integration or by stochastic discrete event simulation on demand. For the latter, the reaction rates are turned into reaction probabilities, following the approach suggested by Gillespie (1976) or its many variants that have been developed since the end of the 1990s (Gibson and Bruck, 2000; Gillespie and Petzold, 2004).

Rather than to implicitly transform a set of reaction equations into a stochastic model, as done by the Gillespie

*Correspondence: R Ewald, Modeling and Simulation Laboratory, University of Rostock, Albert-Einstein-Str, 21, Rostock 18059, Germany. E-mail: roland.ewald@informatik.uni-rostock.de

approach, other approaches use an explicit discrete event representation of their system as a basis for execution. Discrete event modelling directs our view to individual entities and their discrete, often stochastic, dynamics. Variables can be arbitrarily scaled. Sub-systems advance over a continuous-time scale, but in steps of variable size. Discrete event models forgo assumptions about homogeneous structure or behaviour and support an in-detail 'micro' inspection of the modelled system, a view that is increasingly supported by newly available wet-lab techniques. For example, light microscopy of single-cell dynamics has become a reality and enables a researcher to track individual proteins or the concentration of a small number of proteins in real time (Sauro *et al*, 2006).

Unlike continuous systems modelling, discrete modelling lacks a common denominator for model description. Even though general approaches exist, for example, PETRI NETS, STATECHARTS, DEVS, and stochastic π-CALCULUS, these formalisms are less known, differ significantly, and are seldom supported by off-the-shelf software. Biologists have to be explicitly introduced to them—an introduction that is typically reserved to the steadily increasing number of research projects in which computer scientists and biologists work closely together (Priami, 2006). As none of these approaches have been originally developed to describe cell biological systems, the benefit of using them depends on the system under study and the questions to be answered. Revealed deficiencies spurred the development of various extensions, some of which will be shortly discussed within this paper. Starting with a biological example, we will discuss some of the specific requirements in modelling and simulating cell biological systems.

2. Example—the Wnt pathway

In biology, signal transduction pathways are essential processes for converting a signal or stimulus into another. For example, they are involved in gene activation, metabolism, and cell locomotion in all kinds of biological organisms. The canonical Wnt pathway plays a key role in embryogenesis and homoeostatic processes, as well as in the context of some diseases like colorectal cancer (Wodarz and Nusse, 1998; He, 2003; van Es *et al*, 2003; Zheng, 2003; Logan and Nusse, 2004). The pathway is highly conserved in a wide range of different animal species, which additionally underlines its importance.

Many details of the Wnt pathway are still not completely understood. This is especially true for aspects like cell specificity, kind and manner of activation of the different parts of the signalling pathway, as well as time dependence of the activation. In addition to the canonical one, there are also non-canonical variants of Wnt protein-activated signalling pathways. These pathways are still poorly understood. In contrast to them, the main interactions and events of the canonical Wnt pathway are already well known. Under stimulated conditions, a cascade of several biochemical processes results in a cytoplasmic accumulation of β-catenin and subsequent, activation of target genes. Because of its enormous complexity we reduced the already simplified pathway model from Lee *et al* (2003). Our example contains only the basic components and interactions. These are shown in Figure 1 and the corresponding reactions are given in Figure 2.

The initial signal for this pathway is an extra-cellular binding of a Wnt protein to the seven-pass transmembrane receptor Frizzled (Fz). This causes an activation of the cytosolic protein Dishevelled (Dsh), which now can induce the release of glycogen synthase kinase 3β (GSK3) from the β-catenin destruction complex. Besides GSK3, this complex also contains the proteins adenomatous polyposis coli (APC)

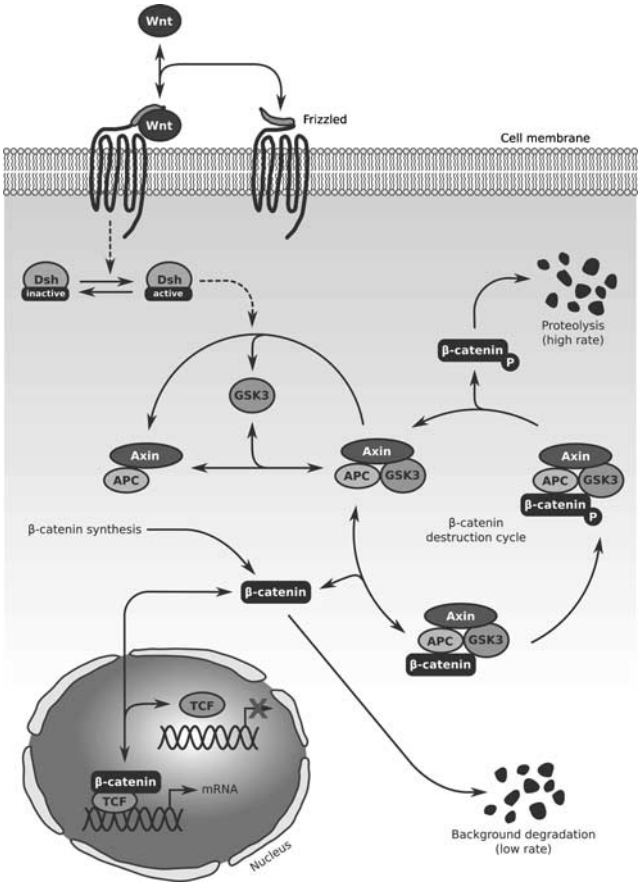


Figure 1 The Wnt signalling pathway: Reactions and complexations of proteins are shown as solid arrows. The double-headed ones denote reversible reactions and the single-headed arrows represent processes that take place only in one direction. Dashed arrows indicate activation of reactions, but the activating species do not participate stoichiometrically in this reactions. Phosphorylated β-catenin is marked by an encircled P. The interacting processes of this system take place in different compartments that are separated by the cell membrane and the nucleus.

$$
\begin{aligned}
\mathrm{Wnt} + \mathrm{Fz} &\Longleftrightarrow \mathrm{Wnt{\cdot}Fz}\\
\mathrm{Dsh}_i &\xrightarrow{\mathrm{Wnt{\cdot}Fz}} \mathrm{Dsh}_a\\
\mathrm{Dsh}_a &\longrightarrow \mathrm{Dsh}_i\\
\mathrm{Axin{\cdot}APC{\cdot}GSK3} &\xrightarrow{\mathrm{Dsh}_a} \mathrm{Axin{\cdot}APC} + \mathrm{GSK3}\\
\mathrm{Axin{\cdot}APC} + \mathrm{GSK3} &\Longleftrightarrow \mathrm{Axin{\cdot}APC{\cdot}GSK3}\\
\emptyset &\longrightarrow \beta\text{-catenin}\\
\beta\text{-catenin} &\longrightarrow \emptyset\\
\beta\text{-catenin} + \mathrm{TCF} + \mathrm{gene} &\Longleftrightarrow \beta\text{-catenin{\cdot}TCF{\cdot}gene}\\
\beta\text{-catenin{\cdot}TCF{\cdot}gene} &\longrightarrow \beta\text{-catenin} + \mathrm{TCF} + \mathrm{gene} + \mathrm{mRNA}\\
\mathrm{Axin{\cdot}APC{\cdot}GSK3} + \beta\text{-catenin} &\Longleftrightarrow \mathrm{Axin{\cdot}APC{\cdot}GSK3{\cdot}}\beta\text{-catenin}\\
\mathrm{Axin{\cdot}APC{\cdot}GSK3{\cdot}}\beta\text{-catenin} &\longrightarrow \mathrm{Axin{\cdot}APC{\cdot}GSK3{\cdot}}\beta\text{-catenin*}\\
\mathrm{Axin{\cdot}APC{\cdot}GSK3{\cdot}}\beta\text{-catenin*} &\longrightarrow \mathrm{Axin{\cdot}APC{\cdot}GSK3} + \beta\text{-catenin*}\\
\beta\text{-catenin*} &\longrightarrow \emptyset
\end{aligned}
$$

Figure 2 Reaction schemes: Irreversible Reactions are written with a single-headed arrow and reversible reactions with a double-headed arrow. Molecular species above arrows activate these reactions, but do not take part stoichiometrically in the reaction scheme (dashed arrows in Figure 1). Dots between two or more species means that this is a complex of these molecules. Phosphorylated β-catenin is marked by an asterisk. β-catenin synthesis/degradation reactions define an empty set ($\emptyset$) as a reactant/product.

and the scaffolding Axin. In absence of a Wnt signal, the complex leads to a continuous reduction of the β-catenin concentration to a very low level. The β-catenin proteolysis is a consequence of its phosphorylation by GSK3 within the destruction complex. The degradation of β-catenin inhibits its translocation into the nucleus and therefore, also the binding to the T-cell-specific transcription factor (TCF). Summarizing this part of the signalling process, without a Wnt signal the gene activity is downregulated, because β-catenin is continuously degraded by the destruction complex. In the presence of Wnt protein, the destruction complex is degraded itself, and so the cytoplasmic concentration of β-catenin increases. The more β-catenin exists in the cytoplasm, the more of it can diffuse into the nucleus and can activate the gene transcription by binding to TCF.

2.1. Different aspects for the modelling of biological systems

During the modelling process of biological systems, various aspects have to be considered, because there are many different ways how the individual components can be modified or interact with each other. For example, proteins can be activated or inactivated by phosphorylation and dephosphorylation, respectively, there are binding equilibria for formations of protein complexes, and in addition to reversible reactions we have also irreversible ones. Another very important aspect of signalling pathways is signal amplification. For example, the activated Dishevelled protein in the canonical Wnt pathway can induce the release of GSK3 from more than one β-catenin destruction complex (Axin/APC/GSK3). So, a single activated Dsh protein leads to a higher amount of destructed Axin/APC/GSK3 complexes and subsequently to a once more higher accumulation of β-catenin in the cytoplasm. Because of stochastic effects in such amplification cascades, an unambiguously stoichiometric relationship is not given. Stochastic effects might also play an important role if only a few entities of one species exist. For example, the Axin concentration in the Wnt signalling pathway is very low in comparison to other players of this pathway (Lee *et al*, 2003). It is one hundred to thousand times lower than the overall β-catenin concentration.

In such cases, continuous models using differential equations do not map the real system very precisely. Here, a discrete event approach could be more feasible for modelling such aspects. Spatial separations of different processes are also very important aspects of biological systems. In the Wnt pathway, for example, the binding of β-catenin to TCF takes place in the nucleus, which distinguishes this compartment from the cytoplasm. So, the concentration of β-catenin in the nucleus is strongly influenced by its diffusion rate through the nuclear pores. Although we do not yet have sufficient spatial data to describe such aspects precisely, current research in different research groups is under way to collect these data and to develop models in which spatial information is included.

3. Discrete event modelling

DEVS (Zeigler, 1984), PETRI NETS (Murata, 1989), STATECHARTS (Harel, 1987), and stochastic π-CALCULUS (Priami, 1995) (in the following also stochastic π) are formal and generally applicable approaches towards discrete event systems modelling. Their use has been explored in Systems Biology, and depending on their success inspired a broader exploitation and the refinement of methods. The latter becomes particularly obvious in the area of process algebras, which have received an increasing attention over the last years.

The formalisms have been developed with a rather different objective in mind. For example, the goal of DEVS has been to combine the functional, network, and hierarchical perspectives in describing systems, and stands in the tradition of general systems theory. In contrast, PETRI NETS and π-CALCULUS have been developed for describing concurrent processes and are best known in the context of Computer Science. Whereas PETRI NETS are particularly used for concurrent systems that are competing for resources, the focus of process algebras is on the continuation and communication of processes. STATECHARTS are also very prominent in Computer Science, albeit in the domain of software development.

3.1. Modelling a biological system

PETRI NETS. PETRI NETS are a visual modelling formalism that offers a natural description of reaction networks. Unlike the usual depiction of reaction networks, PETRI NETS combine the graphical representation with a rigorous formal semantics which, due to the comparatively long tradition of PETRI NETS research in Computer Science, comes along with a huge set of different simulation and analysis tools.

PETRI NETS are directed, bipartite graphs in which nodes are either 'places' (circles) or 'transitions' (rectangles). Circles are only connected to transitions and vice versa. A PETRI NET is marked by placing 'tokens' on places. A transition is enabled if the input places of a transition provide the required number of tokens and the output places have the required capacity for the transition to fire. If a transition fires, the required tokens are removed from the input places and new ones are generated at the output places.

In stochastic PETRI NETS (SPN), exponentially distributed stochastic delays are associated with the transitions (Wilkinson, 2006) and provide the means to describe stochastic processes from molecular biology in a discrete event manner (Goss and Peccoud, 1998). This allows to define a place for each kind of entity in the model, and to describe the reactions between them as transitions (see Figure 3).

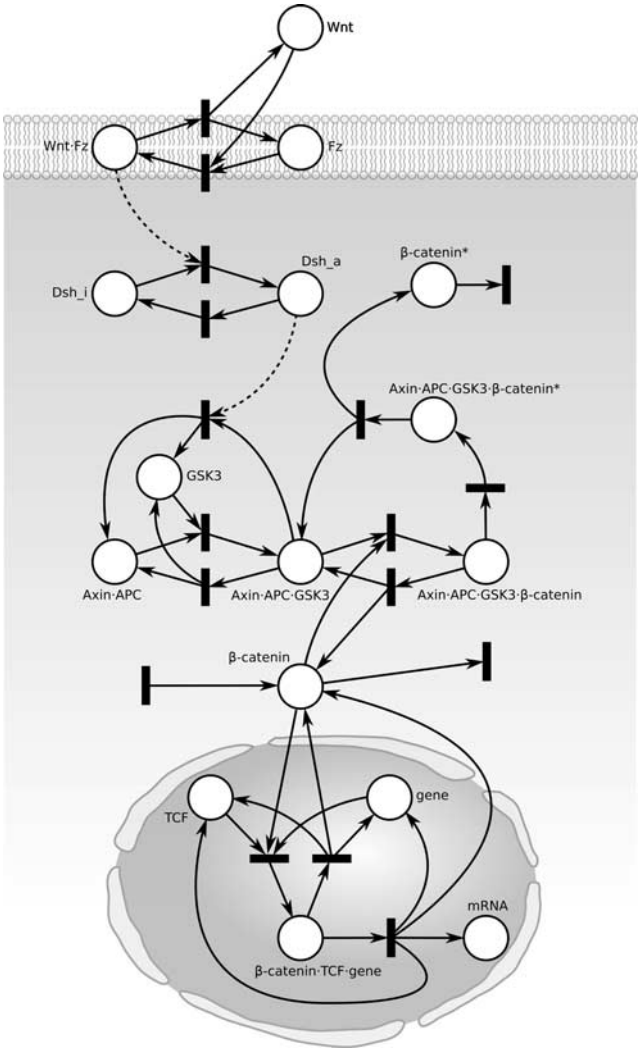


Figure 3 PETRI NET of the Wnt signalling pathway.

π-CALCULUS. The π-CALCULUS was created as a theoretical framework for the study of concurrent computation. Its primitives are *interactions* and *processes*. Similar to PETRI NETS, it has undergone profound theoretical studies. It was first applied in the domain of concurrent programming languages. Although it has been suggested for molecular modelling only recently (Regev and Shapiro, 2002), it meanwhile finds a widespread use in the area of computational biology.

The π-CALCULUS can be understood as a minimal programming language. It is based on *names* and uses a small set of operators to create terms that are referred to as *processes*. The syntax is accompanied by a semantics that specifies the interpretation of processes. There is only one rule of action in the π-CALCULUS. It allows two concurrent processes to interact using a name they share the knowledge of. This name denotes a *channel*, over which one process acts as a sender, while the other acts as a receiver. The transmitted message is again a name, which the receiver henceforth knows and may use in further interactions. Such message-passing allows to describe networks with evolving connectivity. Congruence laws define equality of processes. In an extension, exponentially distributed stochastic delays have been assigned to the interactions, which constituted the stochastic π-CALCULUS (Priami, 1995). Thus, similar to original PETRI NETS, the original π-CALCULUS has not foreseen a stochastic and thus quantitative interpretation, but focused on a qualitative perception of the system. Channels have stochastic firing rates with exponential distribution, so that the non-deterministic choice of the original π-CALCULUS becomes a stochastic race. Some channels are globally known and introduce a macro-perspective. Interaction over a channel may therefore represent a reaction between molecules (ie processes), as thousands of processes might concurrently wish to interact over one channel (ie a binding site).

Processes in stochastic π-CALCULUS might occur in parallel (denoted by the | operator) or in a sequence (denoted by a comma). Additionally, the + operator allows to express non-deterministic choice or a stochastic race in π-CALCULUS and stochastic π-CALCULUS, respectively. Figure 4 shows a stochastic π-CALCULUS model of the Wnt pathway example. All process names start with capital letters and all channel names with a lower case. Each of these rules defines the 'behaviour' of one kind of process. For example, the definition of Wnt as (*new w@delwnt*) (*!wntChannel(w)*,

?w, *Wnt*) tells us the following: first, a *Wnt* process tries to send a newly created private channel *w* over the global channel *wntChannel*. This is expressed by '(*new w@delwnt*) (*!wntChannel(w)*, *[…]*)', where *w@delwnt* means that the newly generated channel is initialized with a specific rate (*delwnt*). After sending the channel *w*, the *Wnt* process waits until it receives something over it (*?w*). When this happens, it goes back to its original state (*Wnt*).

The only kind of process that listens to the global *wntChannel*, and thus can interact with a *Wnt* process, is *Fz*, representing the Frizzled receptor on the cell membrane. An *Fz* process is defined as *?wntChannel(f)*, *WntFz*. This means, it firstly waits for an input on the *wntChannel*, and names the channel it receives *f*. Then, the process is defined by '*WntFz*→(*!dshChannel*, *WntFz*)+(*!f*, *Fz*).', which describes a stochastic race between activating *Dsh* (by sending over the global *dshChannel*) and releasing the binding to the 'docked' *Wnt* process by sending over *f* and going back to *Fz*. Executing *!f* will make the associated *Wnt* channel receive over its private channel, and so it can proceed from *?w* to *Wnt*. The rest of the model works similarly. Note that the β-catenin process (*Beta*) sends two channels over the *betaChannel*, namely *b* and *bp*. These are received by an instance of *AxinApcGsk3* and will then be used by *AxinApcGsk3Beta*, to signal the *Beta* process if it gets released (channel *b*) or phosphorylated (channel *bp*).

STATECHARTS. STATECHARTS have been developed to describe the internal workings of software objects, inspired by finite state machines. Due to their intuitive graphical representation and their powerful yet easy to grasp semantics, STATECHARTS are extremely prominent in the domain of model-based software development, namely as an integral part of the Unified Modelling Language (UML).

In STATECHARTS, active entities are distinguished from passive ones. The latter are properties or behaviour attributed to a STATECHART, becoming part of its state and transition functions. A state reflects a situation in the life of an object during which this object satisfies some condition, performs some action, or waits for some event. Whether this object refers to a molecular species, a reactant, a product, or even a reaction, is a decision of the modeller. Similarly, each of these biological components can be described as part of a state or as a state transition. Thus, STATECHARTS allow to structure our knowledge into different abstraction levels. They can be easily animated and intuitively show what happens in a reactive machine.

An exemplary STATECHART for the Axin/APC complex in the Wnt pathway is depicted in Figure 5. It comprises several states that indicate whether a GSK3 or β-catenin molecule is docking. Other species could be modelled in a similar manner, and the individual entities could then be executed concurrently. A β-catenin entity may either dock or undock from an Axin/APC complex, which is done by sending and receiving messages (see state transitions in Figure 5). Such cases are problematic, since the reaction time of β-catenin depends on the *amount* of Axin/APC entities in the correct state (which is '*Axin*/*APC*/*GSK3*'). Therefore, the required random selection from all individuals in the right state is assumed to be provided by some internal functions (eg 'getAllActDsh()' in Figure 5). To avoid artefacts like this, one could also use STATECHARTS to describe a more abstract model of the Wnt pathway, in which single molecules are represented by messages (see Figure 6).

Note that not all molecules have been modelled as individual entities, some of them are only represented as aspects of another molecule's internal states (eg GSK3, Figure 5). This decreases the complexity of the model, but it also decreases the reusability of such models, for example, in

```
Wnt → (new w@delwnt) (!wntChannel(w) , ?w , Wnt) .
Fz → ?wntChannel(f) , WntFz .
WntFz → (!dshChannel , WntFz) + (!f , Fz) .
Dsh_i → ?dshChannel , Dsh_a.
Dsh_a → (@delay , Dsh_i) + (!dshActiveChannel , Dsh_a).
Gsk3 → (new g@delgsk) (new go@zerodelay) (!gsk3channel(g,go),
(?g+?go), Gsk3).
AxinApc → ?gsk3Channel(g,go) , AxinApcGsk3 .
AxinApcGsk3 → (?dshActiveChannel , !go, AxinApc) + (!g, AxinApc) +
(?betaChannel(b , bp) , AxinApcGsk3Beta) .
AxinApcGsk3Beta → (!b , AxinApcGsk3) + (!bp , AxinApcGsk3BetaP) .
AxinApcGsk3BetaP → @delay , AxinApcGsk3 .
Beta → (new b@delbc) (new bp@delph)(new bt@delbt) (!betaChannel(b, bp) ,
 (?b , Beta) + (?bp, @delay))+(@delay)+(!betaTcfchannel(bt), ?bt, Beta) .
BetaSynth → @delay , (BetaSynth | Beta) .
TcfNoBeta → ?betaTcfChannel(bt),TcfBeta .
TcfBeta → (!bt,TcfNoBeta)+(@delay,(mRNA | TcfBeta)) .
```

Figure 4 Stochastic π model of the Wnt signalling pathway.

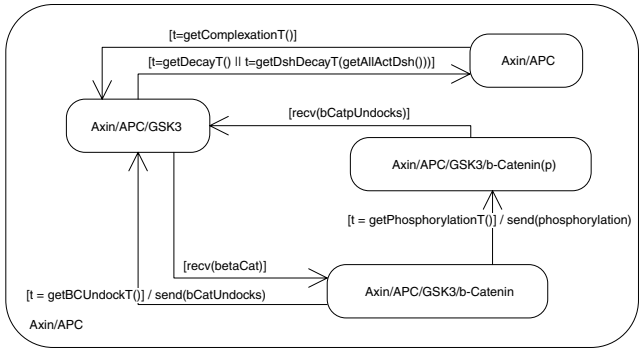


Figure 5 STATECHART of the Axin/APC complex from the Wnt signalling pathway: While this STATECHART represents a complex of two molecules, namely Axin and APC, other STATECHARTS might describe individual molecular entities. On the other hand, the presence of molecules may also be encoded in the state. For example, the presence of GSK3 is apparent in all states but the one in the upper right corner, namely 'Axin/APC'.

case that GSK3 would require an additional state and therefore a representation as a STATECHART on its own.

However, a reuse of the reactive machine Axin/APC would follow a different argumentation line. From a more general view, we defined a reactive machine that is based on a component whose velocity of unbinding or binding with another one depends on a signal. A third component can bind, and this component might fall off or might get phosphorylated. This reactive machine could be reused to describe similar reaction patterns by simply using different parameters, for example, for the reaction rates and the type of components that can bind to it.

DEVS. We wish to conclude our exploration on discrete event modelling formalisms with DEVS, which is rooted in systems theory. It has been developed as a general framework for modelling and simulating dynamic systems and is supported by a large number of simulation engines. Like STATECHARTS, it focuses on *states* and *transitions* between those.

DEVS models are either *coupled*, that is, consisting of different sub-models, or they are *atomic*. The formalism supports the hierarchical modular composition of models. Each model communicates with its environment via its input and output ports. Coupled models are defined by their components and the couplings that exist between those. Coupled models do not have a behaviour of their own. As the formalism shares several similarities to STATECHARTS, DEVS models are often visualized by them, particularly at atomic level. Similarly to STATECHARTS, state transitions of atomic models are either triggered by the arrival of external events or by the flow of time. If an external event arrives, the elapsed time since the last transition has happened is taken into account. To trigger events by the flow of time, each state is associated with a lifetime. The interactions between models are timeless and constrained at the level of the coupled model (by the defined couplings between output and input ports). If an output port is connected to several input ports, messages will be sent simultaneously to all recipients. What happens with the produced output is of no concern to the sending model, its knowledge about its environment ends at its output ports. The DEVS example of the Wnt pathway is shown in Figure 7.

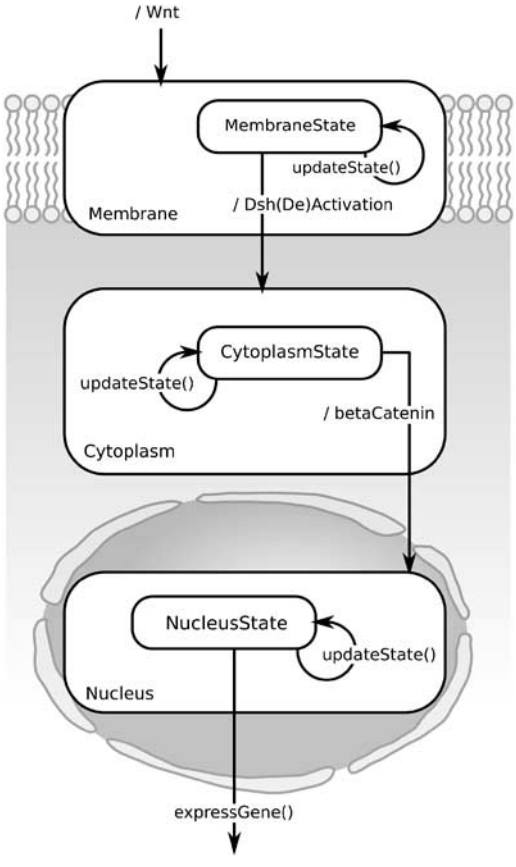


Figure 6 Macro-view STATECHART of the Wnt signalling pathway: To avoid modelling artefacts, one could also regard molecules as messages that are exchanged between different parts of the cell.

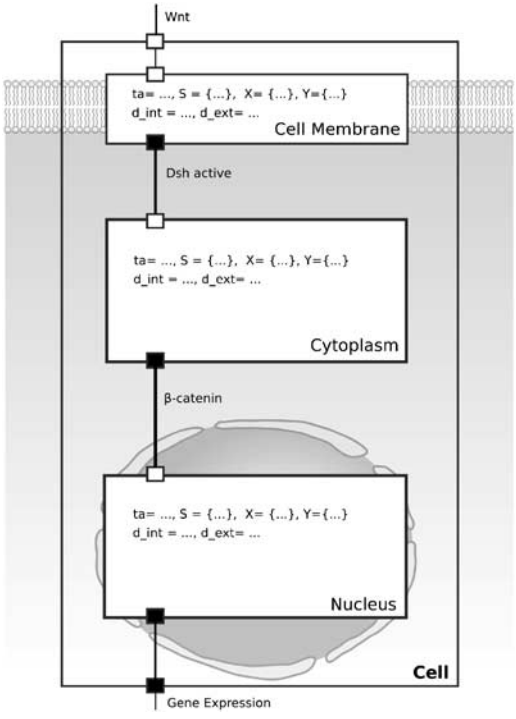


Figure 7 Macro-view Devs model of the Wnt signalling pathway: Similarly to STATECHARTS, one can easily use Devs to describe the macro-view of the signalling pathway. Input and output ports are depicted by white and black boxes. Note that the I/O ports of the coupled model 'Cell' are connected to I/O ports of its sub-models. On the level of the coupled model, input ports are connected to output ports. The internal functions ('ta = …') hint at the formal definition of atomic Devs models (see Zeigler *et al*, 2000).

3.2. Biological aspects

Reactions. Since all biological systems base on chemical reactions, it is essential that these can be modelled intuitively.

Of the presented formalisms, PETRI NETS surely provide the most intuitive way of modelling a reaction network in a discrete event manner at first sight. However, certain phenomena that are most common in biological systems, for example, reversible reactions (see Figure 3), require additional effort, and the circumscription renders describing larger systems by PETRI NETS cumbersome.

In stochastic π, the reacting and interacting entities of the system move into the focus of interest. Reactions are modelled as synchronous communication between individual processes via global channels. Thus, the structure of the reaction network is still visible, but due to the micro-perspective on the involved reactants less so than in PETRI NETS.

Like stochastic π models, STATECHARTS offer a micro-perception rather than reactions, so the involved reactants and products become the focus of interest. In combination with an asynchronous interaction via events, this does not facilitate the representation of chemical reactions that involve several components at once. Although reactants and products can be considered as independent entities, they have to react in synchrony. As can be seen in our Wnt example, this is not straightforward as it contradicts the underlying modelling metaphor. Thus, for a 1:1 interpretation of chemical reactions, STATECHARTS are less suitable than other formalisms like PETRI NETS or process algebras. The same applies to modelling chemical reactions in DEVS.

Biochemical components. If biochemical components like proteins are the focus of interest and shall be perceived as autonomous entities, STATECHARTS offer an intuitive, visual representation that is easy to be understood by biologists and supports a straightforward animation during execution. The same is true for DEVS. Each protein is perceived as a sub-system that reacts to 'environmental perturbations' according to its own state and behaviour rules. In DEVS, the interaction structure between those sub-systems is explicitly defined at the level of the coupled model.

In stochastic π, a protein is a process that moves through different states. States that are associated with a temporal delay are either communicating states or introduced by an explicit delay. Individual biological components can be tracked as in STATECHARTS or DEVS. By being able to generate new channels and to communicate those, stochastic π supports dynamic interfaces of the processes. Thus, the boundary between the system itself and its environment is not as clearly defined as in STATECHARTS and DEVS. However, the dynamics facilitate describing phenomena in biochemical systems, like complexation and polymerization.

In PETRI NETS, the individual components are passive tokens that are consumed and produced by the reactions. Tokens in a place account for the number of molecules in a certain state. The view that PETRI NETS offer is a discrete macro-view rather than a micro-view. However, perceiving the active components with their different states and interactions is one of the attractions of introducing discrete event modelling formalisms into the biological realm, as it allows an in-detail inspection and to trace individual actors. Another important advantage of micro-view models is their ability to cope with molecules that have a large state space. Proteins might have numerous phosphorylation sites, each of which can be activated or not. The state of all phosphorylation sites can be regarded as the state the protein is in: when a protein has n phosphorylation sites, these define a state space of up to 2^n states. Adding a place for all of the various states is clearly inefficient and could lead to huge models (Tolle and Le Novére, 2006).

Signalling. Compared to metabolic pathways, the mechanisms of signalling pathways are more complex. In addition to catalytic reactions, there are also many of complex formation and transportation processes. In contrast to metabolic networks, there is less substance flow in signalling pathways. However, there is much more signal flow instead. This flow can be performed, for example, by a single protein, and signal cascades are often amplified by several magnitudes. In such cascades, a stoichiometric reaction equation is not well defined, but the stochastic effects can be described by all discrete event modelling formalisms. So, all of them are able to manage significantly varying molecule concentrations well.

Signals are often performed by modifications of proteins like phosphorylation or methylation of them, which results in functionally different forms of these proteins. So, here we have the same problem as in the previous section. In PETRI NET and stochastic π models, we need an individual place or process respectively for every state of a protein. This leads to a combinatorial explosion in larger networks. In STATECHARTS and DEVS, this aspect can be modelled more clearly and intuitively.

Spatial information. We distinguish between quantitative and qualitative spatial information. With the latter, we refer to compartmental structures of the cell. For the Wnt pathway, for example, this means distinguishing between the membrane of the cell, the cytoplasm, and the nucleus.

Coupled models in DEVS capture this structuring of a cell, and also in STATECHARTS compartments can be easily reflected. Cellular automata, which are typically used for modelling spatial phenomena, are naturally supported in DEVS (Wainer and Giambiasi, 2001): not only the individual automata, but also the individual neighbourhood has a 1:1 counterpart, due to the explicit couplings in DEVS.

Structuring the space is less obvious in stochastic π. By generating and communicating channels, the name of channels is only locally known. Thereby, a locality of proteins enabling certain processes can be described in stochastic π.

PETRI NETS do not support a representation of space. Spatial effects can be implicitly modelled by introducing delays—a solution also continuous models resort to and which is of course possible in all other discrete event formalisms as well. For example, we could add some places and transitions to model the migration of β-catenin into the nucleus.

None of the formalisms really supports locating proteins or reactions in a cell quantitatively. Rather, they provide an explicit means for structuring the model.

3.3. Extensions and discussion

PETRI NETS. PETRI NETS have been used to qualitatively describe biochemical reaction networks since more than two decades (Fuss, 1987; Schuster *et al*, 2000). Their focus is on structural analysis, for example, identifying biochemical pathways as done in the Krebs cycle (Oliveira *et al*, 2003), or topological analysis, which revealed the central role of specific enzymes in the Glycolysis pathway (Zevedei-Oancea and Schuster, 2003). Higher-level PETRI NETS have also been applied, for example, in Chen and Hofestädt (2003). High-level PETRI NETS extend regular PETRI NETS with respect to hierarchy, time, and data. Our focus has been on stochastic PETRI NETS, which associate transitions with exponentially distributed delays.

PETRI NETS put the emphasis on reactions and their interconnections, rather than on the components of the system to be described. This macro-view translates nicely to continuous formalisms like differential equations, which partly explains their attraction when describing biological systems in a hybrid manner (Chen and Hofestädt, 2003; Matsuno *et al*, 2003). Only few extensions of PETRI NETS, for example, stochastic activity networks (Sanders and Meyer, 2002), support a micro-perception on systems, where individual nets interact via places that represent common resources. Instead, PETRI NETS appear as a straightforward translation of, for example, metabolic reaction networks as they are commonly used to describe biological systems. In PETRI NETS, individual enzymes or the DNA are reduced to passive tokens that are consumed and produced. Populations are modelled by multiple tokens (see section 3.1, PETRI NETS) and arc weights, for example, in TimeNet (http://pdv.cs.tu-berlin.de/~timenet/, accessed 14 February 2007) and Design/CPN (http://www.daimi.au.dk/designCPN/, accessed 14 February 2007).

Most PETRI NETS extensions deviate from the leanness of the original formalism, and hence forgo the analytical benefits that come with a simpler approach. These analytical methods are of central interest (Peleg *et al*, 2005) and it is not surprising that PETRI NETS in Systems Biology sometimes only serve as an intermediate representation, so that the existing PETRI NETS tools and methods can be applied (Peleg *et al*, 2005; Talcott, 2006). For example, the set of input and output places represents all reactants and products being in specific states, the equivalent of a PETRI NET's incidence matrix is the stoichiometric matrix of the corresponding reaction network, and the structure of a PETRI NET reflects largely the biochemical topology. All metabolic or signalling routes that are both stoichiometrically and thermodynamically feasible can therefore be identified by analysing a PETRI NET's structure. Steady state analysis (Voss *et al*, 2003) is supported, as well as comparing reaction networks from different databases (Küffner *et al*, 2000), since discrete event simulation is not the only way to evaluate a model.

π-CALCULUS. As a consequence of focusing on interactions between individual processes, the default perception of stochastic π-CALCULUS is at micro-level. Each individual component can be traced. The calculus is based on the notion of names, which are used to represent both communication channels *and* data. This property allows the interconnection topology of the interacting processes to vary over time. The most prominent advantages of stochastic π-CALCULUS are its simple structure, combined with a strong theoretical foundation and the notion of compositionality. This makes stochastic π-CALCULUS an elegant, minimalistic, and compelling approach to describe cell biological systems. The models can be presented as user-friendly automata (Phillips *et al*, 2006), thereby addressing critiques about the inaccessibility of the formalism.

However, as the formalism has not been designed to describe chemical and biological systems, we also find constructs that appear artificial. The distinction between sender and receiver is not always meaningful (eg in homodimerization, which is the binding process of two identical molecules). Furthermore, one might wish to model more than two molecules forming an active unit, which is possible in other process calculi, for example, PEPA (Calder *et al*, 2006). In addition, means for structuring and reuse are required, so that more complex, spatial models can be described. Inheritance is also not supported, although common behaviour patterns might exist.

Diverse developments, for example, Brane Calculi (Cardelli, 2005), membrane systems (Busi and Zandron, 2006), Bioambients (Regev *et al*, 2004), beta binders (Priami and Quaglia, 2005), or object-oriented extensions (Duchier and Kuttler, 2006) are aimed at addressing these problems. Beta binders provide additional structure to the model, as it is allowed to group processes into binders. It merges the basic features of classical process calculi with the intuition that, in order to model biological entities more closely, simple concurrent processes can be wrapped by borders with

explicit interaction sites. A process in Beta-binders is defined as a box with a proper border and an internal machinery. Boxes make cell compartments explicit, but cannot be nested. The enclosing borders mimic biological membranes and are equipped with typed sites that resemble the motifs of molecules. Biological membranes also inspired the development of other process calculi like Brane Calculi and Bioambients. Again, membranes and compartments are explicitly modelled.

Like PETRI NETS, an important advantage of using process calculi is that, in addition to conventional analysis by simulation, they admit automated verification and falsification of models in the non-stochastic case. Stochastic π models can also be analysed with techniques such as probabilistic model checking (Rutten *et al*, 2004), but the complexity of systems to which these new methods can be applied is still rather limited. Consequently, one of the central challenges for analysis approaches is to support compositionality, independent of whether the models have been defined as SPN or stochastic π processes. Although process calculi are inherently compositional, compositionality is usually only exploited for model description and construction, not analysis. Thus, a replacement of a module with a smaller but provably equivalent one is possible, but there is limited support for compositional quantitative analysis, which enables the derivation of the composed system's properties based on the analysis of individual components (Sauro *et al*, 2006).

STATECHARTS. The view of STATECHARTS seems even more individual-oriented than the perspective of process algebras, because much of the activities are private to an individual STATECHART, and even reactions are individualized. Communication and exchanging events with other STATECHARTS come as a second thought, since activities within a STATECHART are the focus of interest, and less so the interaction with others. The focus on individual STATECHARTS alleviates modelling with respect to questions regarding single entities, for example, in what states an enzyme is observed, and how it reacts to certain events and the flow of time. This perception, together with the asynchronous communication, supports a modular construction of models, and for many systems it seems highly adequate. Whether an individual STATECHART describes a population of molecules or a single one is not determined, because by extending the phases of STATECHARTS by an arbitrary number of arbitrarily scaled variables, different abstraction levels can be modelled easily.

On the other hand, STATECHARTS do not employ an underlying stochastic semantics like stochastic PETRI NETS or the stochastic π-CALCULUS. This leaves the modeller with the responsibility to explicitly define stochastic interactions between STATECHARTS, which introduces artefacts and makes the model more complex. Moreover, communication between STATECHARTS is asynchronous by default. There are operators that synchronize state transitions of two STATECHARTS, but these add additional complexity to the model at hand.

To shift the focus from a single STATECHART to STATECHARTS interaction, life sequence charts have been introduced and combined with STATECHARTS. Life sequence charts form an extension of sequence charts that is aimed at describing the interaction of processes and objects. Life sequence charts distinguish between possible and necessary behaviour, conditions, and progress over time within a chart. Behaviour can be defined both globally, that is, on the level of an entire chart, and locally, that is, when specifying events. This specification can be used as a requirement to check the simulated behaviour, or for animations (a paced type of simulation) in the Play Engine tool (Harel and Marelly, 2003).

DEVS. Much of what has been said about STATECHARTS applies to DEVS as well. Its lack of an implicit stochastic semantics, its focus on individual entities, and an asynchronous communication pattern hamper the modelling of biological systems (see section 3.2, Reactions). In contrast to STATECHARTS, the interaction between individual components is explicitly modelled and the interfaces between systems are clearly described by input and output ports, rather similar to the real-world object-oriented modelling approach. DEVS supports a parallel composition of models via coupling. However, no sequential composition is supported. The general structure of DEVS models is static, a problem that it shares with STATECHARTS and PETRI NETS. The generation of new reactions, new interactions, or new components is therefore not supported by default.

However, since the 1980s, several approaches have been developed to support these variable structure models in DEVS. Recently, also the DEVS extension ρ-DEVS has been proposed, which was motivated by the requirements of molecular biological applications (Uhrmacher *et al*, 2006). ρ-DEVS is based on DYNDEVS, a reflective variant of DEVS that supports dynamic behaviour, composition, and interaction patterns. In ρ-DEVS, dynamic ports and multicouplings are introduced, whose combination allows models to reflect significant state changes to the outside world and enabling or disabling certain interactions at the same time. In addition, a coupled model could be equipped with a high-level model that allows to hold the macro-state of the system and generates the corresponding events, for example, to initiate reactions between individual molecules (Ewald *et al*, 2006).

3.4. Summary: Modelling

Typically, different players in a biological system are translated into models based on the different formalisms as described in Table 1. Whereas in PETRI NETS and Stochastic

π we find a clear correspondence between entities in the biological system and the modelling formalism, modelling is not as straightforward in StateCharts and Devs. For example, individual molecules can be described as entire StateCharts, thereby they are perceived as reactive machines. They can also be encoded within the states, thereby they are interpreted as the subject of manipulation. Thus, StateCharts and Devs allow us to weight the importance of the individual players. However, this has to be decided on a case-by-case basis, and the combination of different views does not necessarily lead to an intuitive model, as the Axin/APC example shows (see Figure 5).

4. Discrete event simulation in systems biology

As biological systems are governed by the laws of chemistry and physics, simulation approaches from these domains are also viable for the simulation of Systems Biology models. Nevertheless, the sheer size of these models often limits their application. This led to numerous abstractions: from the actual physical processes as we know them, described by quantum mechanics etc, over approaches that abstract to entire atoms (*molecular dynamics*), towards approaches that only consider molecules, compartments, or cells (Vaidehi and Goddard III, 2001).

Simulation algorithms for a sub-molecular scale rely on natural laws that are of continuous nature, so these approaches are executed using techniques for continuous simulation (Takahashi *et al*, 2004). On the abstraction level of molecules, one can use approaches that abstract from the natural laws by assuming that the molecules move randomly (ie Brownian Motion). It is quite common to even abstract from single molecules to concentrations of molecules, since one can then describe a system using ODEs, which are relatively simple to compute. However, this deterministic continuous simulation is inadequate for models with small numbers of molecules, where stochastic effects play a role. Consequently, some approaches divide the system into a part that should be simulated with a discrete event approach, and another one that can be simulated continuously (eg Takahashi *et al*, 2004).

The stochastic effects that occur can be expressed by the Chemical Master Equation (CME), which accurately models the system behaviour as a probability distribution of a chemical system's state, depending on the current time. To compute this formula is a fundamental problem when simulating chemical systems, since it is extremely hard to solve analytically in practice.

4.1. Gillespie's exact stochastic simulation

Gillespie (1976, 1977) introduced a stochastic algorithm that simulates a single trajectory of a chemical system. He proofed that this algorithm's outcome exactly samples the chemical master equation, that is, one can arbitrarily approximate the CME by computing more and more trajectories with Gillespie's algorithm.

To compute a trajectory, the algorithm makes a strong assumption: the system under study has to be in thermal equilibrium, that is, all molecules are randomly distributed in a uniform manner. This assumption makes it possible to calculate the so-called *propensity* of each reaction, which is the parameter λ of an exponential distribution. The exponential distribution is then used to model the occurrence of each reaction. The propensity is computed using elementary combinatorics: first, the number of all possible combinations of reaction partners in the solution is calculated. This number is then multiplied with the *stochastic rate* of the reaction, which results from converting the deterministic rate constant, taking into account the volume of the solution and other physical parameters, such as pressure or temperature. The calculation of propensities is common to all variants of Gillespie algorithms.

As an example, the stochastic rate calculation of the reaction 'Axin/APC/GSK3/β-catenin $\rightarrow$ Axin/APC/GSK3/β-catenin*' (see Figure 2) will be given. The protein complex 'Axin/APC/GSK3' will be abbreviated with C for more clarity. The concentration of 'Axin/APC/GSK3/β-catenin',

Table 1 Representation of biological components in the different modelling formalisms

Biological component	*Stochastic Petri Nets*	*Stochastic π*	*StateCharts*	*Devs*
Molecular species	Place	Set of parallel processes	StateChart, state	Atomic model, state
Individual molecule	Token	Process	StateChart, state	Atomic model, state
Reaction	Transition	Channel name	StateChart, state transition	Atomic model, state, coupling
Reactant of reaction	Input place	Process	StateChart, state	Atomic model, state
Product of reaction	Output place	Process	StateChart, state	Atomic model, state
Rate of reaction	Weight of transition	Channel rate	Time delay of state transitions	Time-advance function (*ta*)
Reaction taken place	Transition fires	Synchronous interaction on channel	Event triggers state trans./output	Asynchronous interaction over coupling
State	Markings of places	Inferred by syntax after each reduction step	Global state	Global state

as well as the deterministic rate constant of the reaction, k, were taken from Lee *et al* (2003). As a volume V, a cube with 100 μm edge length is chosen. The concentration has to be multiplied with the Avogadro number n_A and the volume V. This yields the total amount of this type's molecules within the volume.

$$\begin{aligned} C \cdot \beta\text{−catenin} &= 2.02\times10^{-3}\text{nM} \\ k &= 206\,\text{min}^{-1} \\ V &= 100\mu\text{m}\times100\mu\text{m}\times100\mu\text{m} \\ &= 10^{6}\times10^{-18}\text{m}^3 = 10^{-12}\text{m}^3 \\ &= \boldsymbol{10^{-9}}\textbf{L} \end{aligned}$$

$$\begin{aligned} N_{C\cdot\beta-catenin} &= n_A\times2.02\times10^{-3}\text{nM}\times10^{-9}L \\ &= (6.022\times10^{23}\text{mol}^{-1}) \\ &\quad\times(2.02\times10^{-12}\text{M})\times10^{-9}\text{L} \approx \textbf{1217} \end{aligned}$$

The formula that converts the deterministic rate k into a stochastic rate c depends on the order of the reaction (ie the number of reactants). Having only one reactant, $c = k$ holds, that is, the stochastic rate constant equals the deterministic one. Finally, the propensity a of a reaction is calculated by multiplying c with the number of all possible combinations of reactants, which is simply the number of existing complexes, $N_{C\cdot\beta\text{-catenin}}$, in this case. For higher-order reactions, the number of all possible reactions can be deduced using combinatorics. See Wilkinson (2006) for further details.

$$\begin{aligned} c &= k = 206\,\text{min}^{-1} = 3,4\bar{3}s^{-1} \\ a &= N_{C\cdot\beta-catenin}\cdot c = \textbf{4178,36}s^{-1} \end{aligned}$$

Gillespie (1976) introduced two different variants to compute a trajectory based on his approach, namely the Direct Method and the First Reaction Method. Following the direct method, one sums up the propensities of all reactions and uses this sum as the parameter to generate an exponentially distributed random variable, which determines when *any* of the reactions will occur next. Now that the time of the next (reaction) event is known, the reaction that occurs at that time is chosen randomly again. The probability of each reaction to be selected is proportional to its propensity. Finally, the chosen reaction is executed, that is, the number of species in the volume is updated accordingly, and afterwards the propensities have to be updated as well.

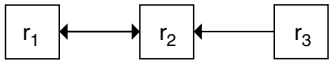
The First Reaction Method, on the other hand, generates the exponentially distributed time of next event for each reaction individually. Then, the reaction with the minimal time of next event, that is, the *first reaction* to be executed, is chosen. The reaction is executed and the algorithm continues with updating the propensities and generating new random times of next event for each reaction.

Although both methods are equivalent, their performance may differ strongly. This is caused by the use of different operations, which might be costly (note that there might be numerous reactions, and the computation of a single trajectory may take millions of iterations): For example, the Direct Reaction method needs to generate only two random numbers per iteration, whereas the First Reaction variant generates $r+1$ random numbers per iteration, r being the number of reactions.

Gibson and Bruck propose some enhancements to Gillespie's original algorithms in Gibson and Bruck (2000), which they integrate in a new (yet equivalent) variant, the Next Reaction Method. Most notably, the new method reduces the time-consuming recalculation of reaction propensities. This is achieved by constructing a directed dependency graph on initialization. When a reaction is executed, not necessarily all of the species' populations have been changed. The dependency graph is used to identify the reactions whose propensity requires an update, and only those propensities are recalculated. The idea bases on the fact that the propensity of a reaction changes if and only if one of its reactants is also a product of the executed reaction (see Figure 8). Furthermore, Gibson and Bruck's approach linearly interpolates the reaction times of all updated reactions, so that the generation of additional random numbers is avoided. The approach does indeed generate only one random number per iteration, to calculate the next reaction time of the executed reaction.

4.2. Tau Leaping

Although the aforementioned approaches are very helpful for small systems and provide an exact stochastic simulation, their performance prohibits an application to systems of a larger scale. This is especially true for systems with concurrent reactions of differing speed (eg gene expression and metabolic reactions): if populations of the metabolites are sufficiently high, many iterations (and propensity updates) are needed without any significant changes in any reaction propensity. This problem is equivalent to the



r_1 : Axin•APC + GSK3 → Axin•APC•GSK3

r_2 : Axin•APC•GSK3 → Axin•APC + GSK3

r_3 : Axin•APC•GSK3·β-catenin* → Axin•APC•GSK3 +β-catenin*

Figure 8 Dependency Graph: While reactions r_1 and r_2 may influence each other, r_3 does only influence r_2. This is because its occurrence does not change the amount of Axin/APC or GSK3 molecules (which influence the propensity of r_1), but that of Axin/APC/GSK3 molecules.

challenge of numerically integrating stiff ordinary differential equation (ODE) systems.

To overcome this problem in the discrete event domain, a technique called Tau Leaping has been introduced by Gillespie *et al* (2001). Tau Leaping approximates the execution of Gillespie's exact approach by leaping forward a time step τ, in which the propensities of all reactions are *approximately* constant. How often each reaction has occurred during this leap can be determined by a Poisson distribution. All reaction occurrences are then executed at once, their propensities are updated, and the algorithm continues by determining the size of the τ leap for the next iteration.

To implement a suitable Tau Leaping method, one has to solve several problems, as outlined in Cao *et al* (2006). When Tau Leaping is applied to systems with small species populations, it might happen that a τ leap results in negative molecule numbers. Theoretically, reactions might occur more often than it would actually be possible, which might render the entire simulation result invalid. This problem can be solved in different ways, for example by replacing the Poisson distribution with a binomial distribution (Tian and Burrage, 2004). Another problem stems from the τ selection mechanism itself. When computing the size of the next leap, the algorithm has to make sure that the assumption of 'nearly constant' propensities holds for all reactions. In the original approach, τ is calculated so that the sum of all propensities is nearly constant, but it is not ensured that smaller propensities remain constant as well. This can introduce a serious error to the simulation result (Cao *et al*, 2006). On the other hand, it is not easy to take the fluctuations of smaller propensities into account, especially without sacrificing the performance by choosing a very small τ. Basically, this trade-off is due to the approximate nature of Tau Leaping: the larger the τ, the faster is the simulation, but the less exact are the results. Quite similar to numerical integration, the research in this area focuses on ways to make Tau Leaping more adaptive, that is, it should choose an appropriate τ, depending on the model at hand. Cao *et al* (2006) tackled this problem with a sophisticated τ selection procedure, but the overall algorithm still includes a fall-back to Gillespie's original approach and a kind of step-size control (cf. step-size control in numerical integration).

4.3. Next sub-volume method

Besides some optimizations regarding the performance, it has also been tried to extend the applicability of Gillespie's approach in the biological domain. In biological systems, gradients of substance concentrations, molecular crowding, and diffusion play important roles in many cases. Additionally, compartmentalization, that is, the isolation of different conditions by membranes, makes these systems also more complex. For example, the Wnt pathway comprises reactions not only near the cell membrane, but also inside the nucleus. Furthermore, the signalling mechanism relies on the diffusion of β-catenin into the nucleus. This effect cannot be easily simulated using Gillespie's original approach, since it presumes a well-mixed solution, that is, uniformly distributed molecules. This assumption is clearly invalid in the cases mentioned above.

Elf and Ehrenberg (2004) propose the Next Sub-volume Method, which introduces spatial information to the Gillespie approach. The original volume is replaced by a number of sub-volumes, and diffusion reactions between those sub-volumes are allowed. The algorithm itself is straightforward: a basic version of the Gillespie algorithm simulates each sub-volume, while a very similar algorithm is used to synchronize the sub-volumes with each other. At first, the time of next event is calculated for each sub-volume, similarly to the First Reaction Method. Then, the sub-volume with the smallest time of next event may execute its next reaction. If this reaction is not a diffusion reaction, it can be executed within the sub-volume and the algorithm re-iterates, after generating a new time of next event for the activated sub-volume. If the executed reaction is a diffusion, a target sub-volume is picked randomly from the activated sub-volume's neighbours and has to update its state. This is illustrated in Figure 9. The Wnt pathway model could, for example, be enhanced with a more detailed model of signalling near the cell membrane: Wnt molecules (black) may diffuse to other extra-cellular sub-volumes, but may also react with a Fz receptor (grey) to form an activated receptor Wnt/Fz (white). Models of the Next Sub-volume Method may comprise thousands of sub-volumes, depending on the required spatial precision.

4.4. Other approaches

In the remainder of this section, we want to discuss the relations of Gillespie's approach to other discrete event

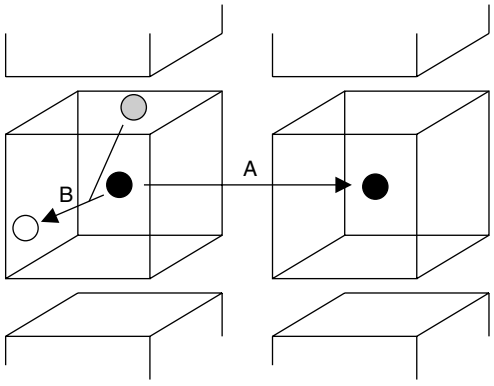


Figure 9 The Next Sub-volume Method: In the sub-volume that was selected to execute the next reaction, there may either be a diffusion into another sub-volume ('A'), or an internal reaction ('B'). If a diffusion occurs, state and reaction propensities of the target sub-volume have to be updated as well.

simulation engines in the domain of Systems Biology. In fact, most of them base on the notion of a continuous-time Markov process. Basically, this is a stochastic process whose behaviour can be determined by its current state (ie it is memoryless) and a matrix of transition probabilities into new states. In the domain of Gillespie algorithms, the state is a tuple containing the population sizes for all species. The relevant part of the (usually huge) probability matrix is then computed using the propensities. In other words, the Gillespie algorithm is an exact computation of the Markov process that underlies the CME (cf. Gillespie, 1976).

But not only the Gillespie algorithm simulates this Markov process; also, other valid, quantitative, and discrete event simulations of a biochemical system approximate the CME in this or in a very similar manner. For example, stochastic process algebras can also be interpreted as specifications for Markov processes, with their state being the current expression, while the channels with their rates as well as the structure of the expression (ie sequences, etc) determine the transition probabilities to the next state (see section 3.1, π-CALCULUS). In modelling and simulation tools like (Priami *et al*, 2001; Phillips and Cardelli, 2004), interactions are scheduled with exponential delays and counting the senders and receivers at one channel. Thereby, they realize the same simulation of the CME as Gillespie's algorithm does. Likewise, stochastic PETRI NETS specify a Markov process, if one regards the assignment of tokens to places as the state, and the transition rates as determinants for the transition probabilities to the next state. Barbuti *et al* (2005) propose a discrete time algorithm that approximates Gillespie's approach, that is, they compute a Markov chain instead of a continuous-time Markov process. This allows the application of existing verification tools, for example, probabilistic model checkers.

The situation is a bit different when simulating DEVS models or STATECHARTS. Both approaches employ a micro-level view on the molecules in the system, instead of simply representing species populations by discrete numbers. That is, they allow a detailed, complex description of molecules, which may even contain additional structure for internal processes (eg the indole channel in the Tryptophan synthase (Degenring *et al*, 2004)). Here, stochasticity can be easily added to describe the behaviour of single model entities, that is, the state transitions of an atomic DEVS model or a STATECHART. Stochastic effects cannot be expressed as easily when it comes to describing the interaction of model entities. This is a serious problem for the simulation of both approaches, as it forces the modeller to circumvent these restrictions by adding artefacts to the model (eg a bulk solution). These artefacts may strongly hamper a simulator's performance, as they often exhibit 'unusual' behaviour towards which the simulator is not optimized (eg they might rely extensively on very expensive operations). This leads us to the conclusion that existing simulators for DEVS and STATECHARTS have to be re-evaluated and possibly redesigned to fit to the needs of Systems Biology applications. Clearly, it is even more advisable to enhance the underlying modelling formalism in this direction, so that the introduction of artefacts is not necessary anymore. First steps towards enhancing DEVS have already been taken (Hunt *et al*, 2005; Uhrmacher *et al*, 2006). Other approaches, the so-called single-particle simulations, aim at combining the micro-view with stochasticity on the simulation level (eg MCell (http://www.mcell.cnl.salk.edu/, accessed 14 February 2007) or StochSim (Le Novére and Shimizu, 2001)). A brief overview and motivation for these methods is given in Tolle and Le Novére (2006). Another survey addressing simulators for spatial models is given in Takahashi *et al* (2005).

5. Conclusion

One has to note that none of the presented approaches was developed with biological systems in mind. Consequently, description languages that are specifically developed for cell biological systems (Kitano *et al*, 2005) are currently under development. Different discrete event modelling formalisms emphasize different aspects of the biological system to be modelled: PETRI NETS emphasize the nature of biochemical reactions, whereas STATECHARTS and DEVS direct the focus of interest towards the involved biological components. Process algebras somehow lie in between, as the global channels refer to the reactions but the processes allow an individual micro-perception of the involved biological components. The view of PETRI NETS is relatively close to the macro-view, as molecular species are typically reduced to passive tokens that are consumed and produced by the transitions. The number of tokens in one place represents the number of molecules being in a specific state. As this global information is also required to calculate the reaction propensities, such networks compute the Gillespie algorithm implicitly. In stochastic π, the counting of molecules being in the required state is done by counting the processes that may send or receive over global channels. The time of reaction events is determined by the stochastic channel rates. In contrast, STATECHARTS focus on individual biological species rather than their interaction. Together with an asynchronous communication via events, their metaphor in modelling seems better suited for describing the activities of entire cells than of chemical or biochemical reactions. Since they support not only parallel but also sequential composition and a modular design of models at different levels of abstraction, STATECHARTS can be a good choice to generate cellular systems.

All in all, each of the presented modelling formalisms has its advantages and drawbacks, and we have to conclude that the right choice for a modelling method largely depends on the focus of the study: What is more important, to model reactions easily and efficiently, or to trace individual

molecular entities with complex states? What tools are available for analysis and simulation? Which formalism provides the best possibilities for reuse when enhancing the model in possible future directions? Which one is the easiest to handle for non-computer scientists? What data are available (eg concerning spatial effects)?

These and many other questions have to be considered, and it is still very much in the hand of the user to make the right choice. To avoid wasting time and resources, it is inevitable to know about the strengths and limitations of the approaches presented here. Although simulators may significantly vary with respect to speed and functionality, the selection of a suitable simulator should not interfere the selection of a suitable modelling formalism. As stated in section 4.4, most discrete event simulators have to solve the same computational problem (of computing a continuous-time Markov process) either implicitly (ie built-in), or explicitly (by simulating model artefacts, eg a bulk solution entity). Hence, chances are that the potential performance of discrete event simulators in Systems Biology is roughly comparable, even for different formalisms. Modelling, on the other hand, is crucial not only for running a simulation in the first place, but also for understanding the underlying biological processes. It should therefore be straightforward and impose as little limits as possible on the modeller.

The interested reader is referred to de Jong (2002) for a general overview of modelling formalisms in Systems Biology, and to Matsuno *et al* (2006) for a survey on PETRI NETS. Several process algebras are introduced in Prandi *et al* (2005), while Priami *et al* (2001) gives an introduction of using stochastic π in Systems Biology. Efroni *et al* (2003) motivates and illuminates the use of STATECHARTS. Further details on different simulation approaches are summarized in Burrage *et al* (2005) and Takahashi *et al* (2005).

Acknowledgements—The research is supported by a grant of the Deutsche Forschungsgemeinschaft (DFG).

References

Barbuti R, Maggiolo-Schettini A, Milazzo P and Troina A (2005). An alternative to Gillespie's algorithm for simulating chemical reactions. In: Gordon P (ed.) *Computational Methods in Systems Biology*. University of Edinburgh.

Burrage K, Burrage P, Hamilton N and Tian T (2005). Compute intensive simulations for cellular models. In: Zomaya AY (ed). *Parallel Computation for Bioinformatics and Computational Biology*. Wiley: New York.

Busi N and Zandron C (2006). Modeling and analysis of biological processes by Mem(Brane) calculi and systems. In: Perrone L, Wieland F, Liu J, Lawson B, Nicol D and Fujimoto R (eds). *Proceedings of the 2006 Winter Simulation Conference*, Vol. **2**, Monterey, CA, USA.

Calder M, Gilmore S and Hillston J (2006). Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA. *Transactions on Computational Systems Biology VII,* Lecture Notes in Computer Science, Vol. **4230**. Springer, pp 1–23.

Cao Y, Gillespie DT and Petzold LR (2006). Efficient step size selection for the tau-leaping simulation method. *J Chem Phys* **124**(4): 044109.

Cardelli L (2005). *Computational Methods in Systems Biology*, Chapter Brane Calculi, Springer: Berlin/Heidelberg, pp 257–278.

Chen M and Hofestädt R (2003). Quantitative Petri Net model of gene regulated metabolic networks in the cell. *In Silico Biology* **3**(0029): 347–365.

Cowan R (2003). Stochastic models for DNA replication. In: Shanbhag D and Rao C (eds). *Stochastic Processes*, Handbook of Statistics, Chapter 4, **21**: 137–166.

de Jong H (2002). Modeling and simulation of genetic regulatory systems: A literature review. *J Comput Biol* **9**(1): 67–103.

Degenring D, Röhl M and Uhrmacher A (2004). Discrete event, multi-level simulation of metabolite channeling. *BioSystems* **75**(1–3): 29–41.

Duchier D and Kuttler C (2006). Biomolecular agents as multi-behavioural concurrent objects. *Electr Notes Theor Comput Sci* **150**(1): 31–49.

Efroni S, Harel D and Cohen I (2003). Towards rigorous comprehension of biological complexity: Modeling, execution and visualization of thymic T cell maturation. *Genome Res* **11**(13): 2485–2497.

Elf J and Ehrenberg M (2004). Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Syst Biol (Stevenage)* **1**(2): 230–236.

Ewald R *et al* (2006). Multi-level modeling with DEVS—A critical inspection and steps towards a feasible approach. *Proceedings of the Winter Simulation Conference WSC 2006, Monterey, California, USA*, December 3–6, 2006.

Fuss H (1987). Simulation of biological systems with PetriNets—Introduction to modelling of distributed systems. In: Moller D (ed). *Advances in System Analysis*. Vieweg, Braunschweig, Wiesbaden, pp 1–12.

Gibson MA and Bruck J (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Chem Phys* **104**: 1876–1889.

Gillespie DT (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* **22**: 402–434.

Gillespie DT (1977). Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* **81**(25): 2340–2361.

Gillespie DT (2001). Approximate accelerated stochastic simulation of chemically reacting system. *J Chem Phys* **115**: 1716–1733.

Gillespie DT and Petzold LR (2004). Improved leap-size selection for accelerated stochastic simulation. *J Chem Phys* **119**(16): 8229–8234.

Goss P and Peccoud J (1998). Biochemistry quantitative modeling of stochastic systems in molecular biology by using stochastic Petri Nets. *Proc Nat Acad Sci* **95**: 6750–6755.

Harel D (1987). Statecharts: a visual formalism for complex systems. *Sci Comput Program* **8**(3): 231–274.

Harel D and Marelly R (2003). *Come, Let's Play: Scenario-Based Programming Using LSCs and the Play-Engine*. Springer: Berlin.

He X (2003). A Wnt-Wnt situation. *Dev Cell* **4**(6): 791–797.

Hunt CA, Ropella GEP, Roberts MS and Yan L (2005). Biomimetic in Silico devices. In: Danos V and Schächter V (eds). *Computational Methods in Systems Biology*, Lecture Notes in Computer Science, Vol. 3082, Springer: Berlin, pp 34–42.

Kitano H (2002). Systems biology: a brief overview. *Science* **295**(5560): 1662–1664.

Kitano H, Funahashi A, Matsuoka Y and Oda K (2005). Using process diagrams for the graphical representation of biological networks. *Nat Biotechnol* **23**(8): 961–966.

Küffner R, Zimmer R and Lengauer T (2000). Pathway analysis in metabolic databases via differential metabolic display (DMD). *Bioinformatics* **16**(9): 825–836.

Kuo D and Keasling JD (1996). A Monte Carlo simulation of plasmid replication during the bacterial division cycle. *Biotechnol Bioeng* **52**(6): 633–647.

Le Novére N and Shimizu TS (2001). STOCHSIM: Modelling of stochastic biomolecular processes. *Bioinformatics* **17**(6): 575–576.

Lee E *et al* (2003). The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. *PLoS Biol* **1**(1): 116–132.

Logan CY and Nusse R (2004). The WNT signaling pathway in development and disease. *Annu Rev Cell Dev Biol* **20**: 781–810.

Matsuno H *et al* (2003). Towards biopathway modeling and simulation. In: van der Aalst W and Best E (eds). *Applications and Theory of Petri Nets*, Lecture Notes in Computer Science, Springer, **2679**: 3–22.

Matsuno H, Li C and Miyano S (2006). Petri Net based descriptions for systematic understanding of biological pathways. *IEICE Trans Fundam Electron Commun Comput Sci* **E89-A**(11): 3166–3174.

Murata T (1989). Petri Nets: Properties, analysis and applications. *Proc IEEE* **77**(4): 541–574.

Oliveira JS *et al* (2003). A computational model for the identification of biochemical pathways in the Krebs cycle. *J Comput Biol* **10**: 57–82.

Peleg M, Rubin D and Altman R (2005). Using Petri Net tools to study properties and dynamics of biological systems. *J Am Med Inform Assoc* **12**(2): 181–189.

Phillips A and Cardelli L (2004). A correct abstract machine for the stochastic Pi-calculus. In: *BioConcur 2004*. Electronic Notes in Theoretical Computer Science, Elsevier, to appear.

Phillips A, Cardelli L and Castagna G (2006). A Graphical Representation for Biological Processes in the Stochastic pi-Calculus In: Corrado P, Anna I, Bud M and Hanne RN. (eds) *Transactions on Computational Systems Biology VII*, Lecture Notes in Computer Science, Vol. **4230**. Springer, pp 123–152.

Prandi D, Priami C and Quaglia P (2005). Process calculi in a biological context. *Bull EATCS* **85**: 53–69.

Priami C (1995). Stochastic p-Calculus. *Comput J* **38**(7): 578–589.

Priami C (ed). (2006). *Computational Methods in Systems Biology: International Conference, CMSB 2006, Trento, Italy, October18–19, 2006, Proceedings*, Lecture Notes in Computer Science, Springer.

Priami C and Quaglia P (2005). Beta binders for biological interactions. In: Danos V and Schächter V (eds). *Proceedings of the 2nd International Workshop on Computational Methods in Systems Biology, CMSB'04*. Springer, **3082**: 21–34.

Priami C, Regev A, Shapiro E and Silverman W (2001). Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Inform Process Lett* **80**: 25–31.

Puchulka J and Kierzek AM (2004). Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. *Biophys J* **86**(3): 1357–1372.

Ramsey S, Orrell D and Bolouri H (2005). Dizzy: Stochastic simulation of large scale genetic regulatory networks. *J Bioinform Comput Biol* **3**(2): 415–436.

Regev A and Shapiro E (2002). Cells as computation. *Nature* **419**(6905): 343.

Regev A, *et al* (2004). BioAmbi-ents: An abstraction for biological compartments. *Theoret Comput Sci* **325**(1): 141–167.

Rutten J, Kwiatkowska M, Norman G and Parker D (2004). Mathematical techniques for analyzing concurrent and probabilistic systems. In: Panangaden P and van Breugel F (eds). *CRM Monograph Series* Vol. 23 American Mathematical Society: Providence, RI.

Sanders WH and Meyer JF (2002). Stochastic activity networks: Formal definitions and concepts. In: *Lectures on Formal Methods and Performance Analysis: First EEF/Euro Summer School on Trends in Computer Science*. Springer-Verlag, New York, Inc.: New York, NY, USA, pp 315–343.

Sauro H *et al* (2006). Challenges for modeling and simulation methods in systems biology. In: Perrone L, Wieland F, Liu J, Lawson B, Nicol D and Fujimoto R (eds). *Proceedings of 2006 Winter Simulation Conference*. Monterey, California, USA.

Schuster S *et al* (2000). Structural analysis of metabolic networks: elementary flux modes, analogy to Petri Nets, and application to Mycoplasma Pneumoniae. In: *German Conference on Bioinformatics*. Heidelberg, Germany, pp 115–120.

Takahashi K, Kaizu K, Hu B and Tomita M (2004). A multi-algorithm, multi-timescale method for cell simulation. *Bioinformatics* **20**(4): 538–546.

Takahashi K, Nanda S, Arjunan V and Tomita M (2005). Space in systems biology of signaling pathways: Towards intracellular molecular crowding in silico. *FEBS Lett* **579**(8): 1783–1788.

Takahashi K *et al* (2002). Computational challenges in cell simulation. *IEEE Intell Syst* **17**(5): 64–71.

Talcott C (2006). Pathway logic: A logical approach to modeling cellular processes. In: Perrone L, Wieland F, Liu J, Lawson B, Nicol D and Fujimoto R (eds). *Proceedings of the 2006 Winter Simulation Conference*, Monterey, California, USA.

Tian T and Burrage K (2004). Binomial leap methods for simulating stochastic chemical kinetics. *J Chem Phys* **121**(21): 10356–10364.

Tolle DP and Le Novére N (2006). Particle-based stochastic simulation in systems biology. *Curr Bioinformatics* **1**: 315–320.

Uhrmacher AM, Himmelspach J, Röhl M and Ewald R (2006). Introducing variable ports and multi-couplings for cell biological modeling in DEVS. In: *Proceedings of the 2006 Winter Simulation Conference*. Monterey, California, USA, pp 832–840.

Vaidehi N and Goddard III WA (2001). Atomic-level simulation and modeling of biomacromolecules. In: Bower JM and Bolouri H (eds). *Computational modeling of genetic and biochemical networks*, pp 161–188, MIT Press: Cambridge, MA.

van Es J, Barker N and Clevers H (2003). You Wnt some, you lose some: Onco-genes in the Wnt signaling pathway. *Curr Opin Genet Dev* **13**(1): 28–33.

Van Gend K and Kummer U (2001). STODE—Automatic Stochastic Simulation of Systems Described by Differential Equations. In: Yi T-M, Hucka M, Morohasi M and Kitano H (eds). *Proceedings of the 2nd International Conference on Systems Biology*. Omnipress, Madison, USA, pp 326–333.

Voss K, Heiner M and Koch I (2003). Steady state analysis of metabolic pathways using Petri Nets. *In Silico Biol* **3**: 367–387.

Wainer G and Giambiasi N (2001). Application of the Cell-DEVS Paradigm for cell spaces modelling and simulation. *Simulation* **76**(1): 367–387.

Wilkinson DJ (2006). *Stochastic Modelling for Systems Biology*. CRC Press: Boca Raton, FL.

Wodarz A and Nusse R (1998). Mechanisms of Wnt signaling in development. *Annu Rev Cell Dev Biol* **14**: 59–88.

Zeigler B (1984). *Multifacetted Modelling and Discrete Event Simulation*. Academic Press: London.

Zeigler BP, Praehofer H and Kim TG (2000). *Theory of Modeling and Simulation*. Academic Press: London.

Zevedei-Oancea I and Schuster S (2003). Topological analysis of metabolic networks based on Petri Net theory. *In Silico Biol* **3**: 29.

Zheng Y (2003). A second Canon. Functions and mechanisms of β-Catenin-independent Wnt signaling. *Dev Cell* **5**(3): 367–377.