Journal of Universal Computer Science, vol. 16, no. 9 (2010), 1152-1175
submitted: 28/10/09, accepted: 15/4/10, appeared: 1/5/10 © *J.UCS*

A Multidisciplinary Survey of Computational Techniques for the Modelling, Simulation and Analysis of Biochemical Networks

James Decraene
(Nanyang Technological University, Singapore
jdecraene@ntu.edu.sg)

Thomas Hinze
(Friedrich-Schiller-Universität Jena, Germany
thomas.hinze@uni-jena.de)

Abstract: All processes of life are controlled by networks of interacting biochemical components. The purpose of modelling these networks is manifold. From a theoretical point of view it allows the exploration of network structures and dynamics, to find emergent properties or to explain the organisation and evolution of networks. From a practical point of view, *in silico* experiments can be performed that would be very expensive or impossible to achieve in the laboratory, such as hypothesis-testing with regards to knock-out experiments or overexpression, or checking the validity of a proposed molecular mechanism. The literature on modelling biochemical networks is growing rapidly and the motivations behind different modelling techniques are sometimes quite distant from each other. To clarify the current context, we review several of the most popular methods and outline the strengths and weaknesses of deterministic, stochastic, probabilistic, algebraic and agent-based approaches. We then present a comparison table which allows one to identify easily key attributes for each approach such as: the granularity of representation or formulation of temporal and spatial behaviour. We describe how through the use of heterogeneous and bridging tools, it is possible to unify and exploit desirable features found in differing modelling techniques. This paper provides a comprehensive survey of the multidisciplinary area of biochemical networks modelling. By increasing the awareness of multiple complementary modelling approaches, we aim at offering a more comprehensive understanding of biochemical networks.

Key Words: biochemical networks, modelling, simulation, analysis, systems biology.
Category: A.1, I.6.4, I.6.5 , J.3

1 Introduction

The evaluation of biochemical networks is an ever advancing challenge, which requires the use of the latest modelling techniques to capture system dynamics and properties. State of the art modelling techniques are drawn from a highly diverse range of different research areas, e.g., mathematics, computer science, statistics, etc. As a result the selection of appropriate modelling techniques for a specific research area has become an arduous task.

When selecting a satisfactory modelling technique, a number of considerations are taken into account. These can be divided into two key points of view:

1. User considerations: limited empirical data, knowledge and familiarity with existing modelling techniques.

2. Modelling considerations: biological accuracy, range of applications, computational complexity.

In the development of systems biology, a variety of modelling techniques for biological reaction networks have been established in recent years [Alon, 2007]. Inspired by different methodologies, five fundamental concepts have emerged and are identified as follows:

– *Deterministic*: Chemical reactions are approximated as continuous deterministic processes at the macroscopic/system level. The system's variable states are uniquely determined by the pre-specified parameters describing the reactions (e.g., molecular concentration, reaction rates, etc.) and initial states of these variables. Given an initial set of pre-specified parameters, deterministic models enable one to monitor, predict and describe the dynamics of the system over time and/or space. Examples of deterministic modelling techniques include: ordinary/partial differential equations [Zwillinger (Ed.), 1992, Polyanin and Zaitsev, 2003, Eungdamrong and Iyengar, 2004, Huang and Ferrell, 1996], Michaelis-Menten models [Heinrich and Schuster, 1996] and power-law models [Vera et al., 2007].

– *Stochastic*: In contrast with deterministic approaches, stochastic models explicitly account for the uncertainty that is involved in molecular processes. The system's variable states are determined by the pre-specified system's parameters and through the use of random variables. By addressing randomness or variability, stochastic models provide a more detailed representation of the system's potential dynamics (and not only the average behaviour as in deterministic approaches). Multiple executions of a stochastic model generate unique (from one another) dynamics/observations. The latter can be used to estimate probability distributions of the system's potential states (assisting in the construction of probabilistic models, see below). Examples of stochastic modelling techniques include: Markov chains [Gomez et al., 2001] and chemical master equations [Gillespie, 2001].

– *Probabilistic*: Here, the description of stochastic processes/data is addressed in terms of probability. Probabilistic modelling techniques are deterministic approaches which may infer probabilistic relationships between molecular species/system's states from empirical observations. In contrast with stochastic approaches, a probabilistic model is a statistical inference and description technique which does not represent the underlying stochastic molecular mechanics. Given the initial states of the molecular species, these approaches provide a probability-based description of the system's states. The predictive power of these techniques relies on the probabilistic distributions inferred by the model upon a range of *in vivo/silico* experimental observations (i.e., the training set). An example of probability modelling technique include: Bayesian networks [Sachs et al., 2002] and hidden Markov models [Goutsias, 2006].

– *Algebraic*: Modelling discrete characteristics of chemical reaction networks is principally achieved with algebraic approaches. A common basic assumption for these approaches is a finite or recursive enumerable number of elementary objects. Each object is considered as the smallest unit that can be processed by the system model. In particular, a definition of objects determines the granularity and abstraction level of corresponding models (hierarchically composed of objects, classes of objects, and temporal interaction rules). Both biomolecules and processes can form these objects. Interaction between these objects is usually specified by a relationship between system configurations. The whole system description is based on discrete transitions. This allows structural and comparative analysis of both system composition and behaviour, independent of numerical simulation results. Examples of algebraic modelling techniques include: P-systems [Paun, 2002, Paun et al., 2006], broadcast language [Holland, 1992], Alchemy [Fontana and Buss, 1994], Boolean networks [Genoud and Metraux, 1999], π-calculus [Regev et al., 2001] and Petri nets [Reddy et al., 1993].

– *Agent-based*: Agent-based models (ABMs) extend the algebraic framework by introducing richer features in the computational units (i.e., agents). ABMs are commonly implemented with Object-Oriented programming environments in which agents are instantiations of object classes. The latter is a collection of properties (e.g., size, location, concentration, etc.) and methods (e.g., move, die, react, etc.). Agent-based simulations typically involve a larger number of molecular and/or cellular agents which are executed in a concurrent or pseudo-concurrent manner. Each of these agents possess its own distinct state variables, can be dynamically created/deleted and is capable of interacting with the other agents. The agents' computational methods may include stochastic processes resulting in a stochastic behaviour at the system level. Examples of agent-based modelling techniques include: Stochsim [Le Novère and Shimizu, 2001], Cellulat [Gonzalez et al., 2003] and AgentCell [Emonet et al., 2005]. A review of agent-based techniques is given by [Chavali et al., 2008].

Deterministic and stochastic approaches are the most frequently employed and studied approaches in the field, whereas the attention given to the use of probabilistic, algebraic and agent-based approaches is more recent but rapidly growing.

In this paper we review a number of prominent modelling techniques and examine the individual attributes of each modelling technique. Following on from this we construct a model comparison table. This paper does not attempt to nominate a single most applicable modelling technique, but rather to illuminate the decision process of selecting modelling techniques. Moreover, computational inference methods (i.e., techniques employed to infer the network structure from experimental data) are not addressed in this paper. This is indeed beyond the scope of this paper, nevertheless the reader may find further details in [Soinov et al., 2003, Laubenbacher and Stigler, 2004, Li et al., 2006, Ponzoni et al., 2007].

2 Principles of biochemical networks *in vivo*

As opposed to engineered networks (e.g., electronic circuits) whose topologies can be easily traced, biochemical network connections are invisible. The circuitry of these natural networks is identified through interactions between their molecules.

Biochemical reaction networks found in pro- and eukaryotic cells represent processes from which higher level properties of life are composed. Despite their high degree of complexity and interdependency, they are hierarchically arranged in modular structures of unexpected order.

A strong division of tasks, predefined reactions and transduction pathways as well as an efficient share of resources characterise biochemical networks. Mainly based on proteins as information carrier with high variability in structure, the range of interconnected reaction processes implies the function of a cell and its subunits. Three essential types of biochemical networks *in vivo* can be distinguished: metabolic networks, cell signalling networks (CSN), and gene regulatory (GRN) networks [Alberts et al., 2003]. Metabolism consists of coupled enzymatically catalysed reactions at a minimum level of free energy. This provides conservative functions for the organism. CSNs perform internal and external information processing in concert with GRNs that control protein synthesis. Slight malfunctions or perturbations within these fine-grained and sensitive network structures can have life-threatening consequences. Modelling, analysing and simulating these networks assist us in understanding and prediction of these complex events.

Proteins form central functional elements of the cell. For instance, they subsume the enzymes, hormones, factors, receptors, messengers, and subsidiary substances of which the cell is composed. Therefore, CSNs and GRNs, as control systems for protein generation based on both inherited genetic data and envi-

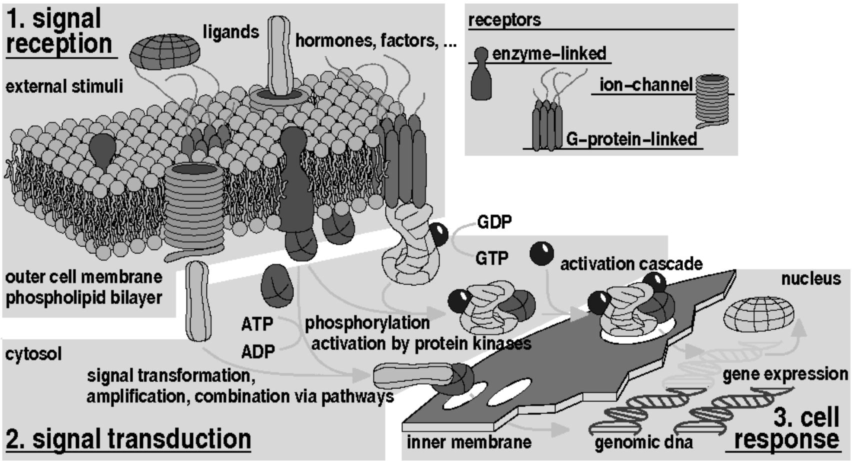

Figure 1: Biological principle of signalling in eukaryotic cells: from arriving stimuli to specific cell response.

ronmental influences, play a major role. In cell signalling, here exemplified by eukaryotic cells [Krauss, 2003], three main steps can be identified (Figure 1):

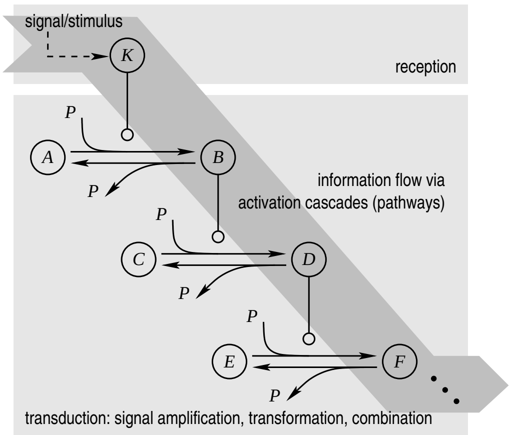

Figure 2: Information flow in CSNs via activation cascades

1. Signal reception: External signals arrive from other cells, from the environment, or from the cell's own feedback loops. These stimuli are encoded either by proteins (like second messenger hormones, growth factors), auxiliary

substances (like ions, ligands), or by physical conditions (e.g., light). They reach specific receptors embedded in the outer cell membrane. Ion channels transmit the signal by transporting substances into the inner cell; whereas enzyme-linked and G-protein-linked receptors transmit the signal simply by changing their conformation.

2. Signal transduction: Messenger proteins, originally bound to these receptors at the inner membrane face, are then emitted into the cytosol. Here, they initiate activation cascades for further signal transduction, evaluation, combination, and amplification, as illustrated in Figure 2. Activation of enzyme messengers occurs by stepwise addition of phosphates from adenosintriphosphate (ATP) to specific binding sites of messenger proteins. Alternatively, G-protein messengers bind to guanosindiphosphate (GDP). These processes can be accompanied by forming specific protein complexes.

3. Cell response: The resulting biomolecules then enter the nucleus where they can effect a specific gene expression controlled by a GRN, thus producing the cell response to the primary signal. The intensity of gene expression is determined by transcription factors. They act as promoters or repressors controlling the amount of mRNA transcribed from genetic DNA. Subsequent translations lead to the final protein. Typical biochemical networks can contain interactions between several hundred proteins including intermediate states and complexes.

In the next Section, our multidisciplinary survey of computational techniques for the modelling, simulation and analysis of biochemical networks is provided.

3 Survey of modelling approaches

We review a selection of modelling techniques used in the study of biochemical networks: differential equations, Markov chains, chemical master equations, Bayesian networks, Term Rewriting Systems, Petri nets, π-calculus, Cellulat and Agent-based Learning Classifier Systems. We then present the Systems Biology Markup Language (SBML) and CellML which allow one to specify and disseminate biochemical network models using a standardised language. These markup languages also permit the migration of reaction network models between differing modelling approaches.

3.1 Differential equations

Chemical reactions are approximated as continuous deterministic processes at the macroscopic level. Differential equations provide a global understanding of a system and are commonly employed to model chemical reaction networks [Zwillinger (Ed.), 1992, Polyanin and Zaitsev, 2003, Eungdamrong and Iyengar, 2004, Huang and Ferrell, 1996]. Given an initial set of pre-specified properties describing the reactions (e.g., molecular concentration, reaction rates, etc.), this modelling approach enables one to monitor, predict and describe the dynamics of the system over time and/or space.

Here, state variables represent the concentrations of molecular species occurring in a well-stirred reactor with no in/out-flows. The following equation governs the dynamics of each species S whose rate of change in concentration $[S]$ depends on the production and consumption rates v_p and v_c:

$$\frac{d[S](t)}{dt} = v_p([S](t)) - v_c([S](t)). \tag{1}$$

In mass-action kinetics, these rates result from the reactant concentrations, their stoichiometric factors $a_{i,j} \in \mathbb{N}$ (reactants), $b_{i,j} \in \mathbb{N}$ (products) and kinetic constants $k_j \in \mathbb{R}_+$ assigned to each reaction quantifying its velocity. For a reaction system with a total number of n species and r reactions

$$\begin{array}{c}
a_{11}S_1 + a_{12}S_2 + \ldots + a_{1n}S_n \xrightarrow{k_1} b_{11}S_1 + b_{12}S_2 + \ldots + b_{1n}S_n \\
a_{21}S_1 + a_{22}S_2 + \ldots + a_{2n}S_n \xrightarrow{k_2} b_{21}S_1 + b_{22}S_2 + \ldots + b_{2n}S_n \\
\vdots \\
a_{r1}S_1 + a_{r2}S_2 + \ldots + a_{rn}S_n \xrightarrow{k_r} b_{r1}S_1 + b_{r2}S_2 + \ldots + b_{rn}S_n
\end{array}$$

the corresponding ordinary differential equations (ODEs) read:

$$\frac{\mathrm{d}\,[S_i]}{\mathrm{d}\,t} = \sum_{j=1}^{r} \left(k_j \cdot (b_{ji} - a_{ji}) \prod_{h=1}^{n} [S_h]^{a_{jr}} \right)$$

In order to obtain a concrete trajectory, all initial concentrations $[S_i](0) \in \mathbb{R}_+$, $i = 1, \ldots, n$ have to be specified. Solving this ODE system together with given initial values allows us to describe the temporal behaviour of the reaction system [Dittrich et al., 2001].

Reaction-diffusion models take into account the spatial location of molecules and allow species concentrations in different spatial locations to vary continuously. These models are specified with sets of Partial Differential Equations (PDEs) [Fritz, 1982]. Solutions to PDEs derived from reaction-diffusion models provide an approximation of the species concentrations as a function $[S](t, x)$ of both time t and space x:

$$\frac{\partial [S](t,x)}{\partial t} = D \frac{\partial^2 [S](t,x)}{\partial x^2} - v([S](t,x)) \frac{\partial [S](t,x)}{\partial x} + v_p([S](t,x)) - v_c([S](t,x)) \tag{2}$$

Equation 2 is an example PDE where the variables and functions represent: $[S]$ concentration of species S, $D \in \mathbb{R}_+$ diffusion coefficient, $v([S](t, x))$ convective velocity, and $v_p([S](t, x)), v_c([S](t, x))$ production and consumption rates.

Differential equations (especially ODEs) are the most commonly employed techniques to model biochemical systems due to their strong establishment in the sciences. Nevertheless using these methods (particularly PDEs) may also represent a significant mathematical challenge when attempting to solve large systems of non-linear differential equations. Moreover, it has been argued that the main challenge of this approach is the limited ability to describe biochemical systems with low species concentrations [Fontana and Buss, 1996]. Chemical kinetic models specify the cell with limited structural descriptions. Biological systems are made of collections of objects whose identities are maintained and continuously evolve. These evolving properties may include the activation state, concentration, or the location.

Since analytic solutions of ODEs can be obtained only in few cases, numerical solutions are commonly employed, predominantly the higher order Runge-Kutta approach characterised by rapid convergence and numerical stability [Atkinson et al., 2009]. The approach is based on discretisation of the time interval and iterative adaptation of the species concentrations. Since each species induces one specific ODE, the computational complexity grows linearly with the number of species.

3.2 Markov chains

Another method to examine biochemical systems is to express them as Markov chains [Gomez et al., 2001], in which the state of the chain represents either approximations or exact number of the molecules present. Reactions are modelled as transitions between these states. The system is memoryless ("Markovian") since the future development only depends on the present, not on the past. Therefore, the term Markov chain denotes time-discrete systems which are defined as a sequence of random variables $X_1, X_2, X_3, ...$ with the Markov property, i.e., $P(X_{t+1} = x|X_t = x_t, X_{t-1} = x_{t-1}, ..., X_1 = x_1) = P(X_{t+1} = x|X_t = x_t)$.

Provided there is no feedback in the system, the analysis of Markov chains is well developed, and the steady-state probability distribution of the process can be derived. Feedback, which is an inherent feature of many reaction networks, poses problems for analysis since a steady-state distribution of the system does not have to exist in this case.

Many straightforward, yet interesting simulation techniques which utilise the Markov property are based on explicit collisions between randomly selected molecules. This technique has the advantage of being easy to implement in a non-spatial case, and yet simple to extend to spatial simulations. A representative example of this type of algorithm is given by StochSim [Le Novère and Shimizu, 2001].

3.3 Chemical master equation

Where the model's time is continuous rather than discrete, the Markov chain is replaced by a "continuous-time Markov process". Here, the system again has a finite, discrete set of states, but now a continuous time index t exists. For simplicity, we focus on the case in which each state is given by the number of molecules per molecular species (i.e., a vector $x \in \mathbb{N}^k$). At any given point in time, the system occupies each state with a certain probability, yielding a probability distribution over all the states. The Chemical Master Equation (CME) provides a means to describe the temporal change of this distribution exactly for the case of a well-stirred and homogeneous reactor space [Van Kampen, 2007]. Since chemical systems can be considered as Markovian, the CME approach is a special case of the continuous-time Markov chains.

[Gillespie, 1976, Gillespie et al., 1977] proposed two precise "Stochastic Simulation Algorithms" (SSA) to simulate instances of the random process defined by the CME. These algorithms are widely used in the stochastic simulation of biochemical reactions [Meng, 2004] due to their significant efficiency in terms of computational cost. The principal factors in SSAs are reaction propensities f_μ, i.e., the likelihood of a reaction μ to occur in the next (small) time step dt. These are computed from the mesoscopic rate constants and the number of molecules available as substrates to the reaction. From these, the next reaction and the time for that reaction have to be decided. This is done by using two random numbers. From the CME, it can be shown that the probability density function for reaction μ to occur as the next reaction after time τ is $P(\mu, \tau) = f_\mu exp(-\tau \sum_j f_j)$, which is the basic equation SSAs are built on.

Gillespie's original work has been extended several times, most notably by the "Next Reaction Method" [Gibson and Bruck, 2000]. This reduces the complexity from linear to logarithmic time in the number of reactions. Another technique is given by the "tau-leap methods" [Gillespie, 2001, Chatterjee and Vlachos, 2005], which approximates the exact solutions obtained from SSAs. For larger numbers of molecules and reactions, however, these algorithms still suffer from high computational requirements. [Bernstein, 2005] extended the Gillespie algorithm to reaction-diffusion equations by dividing the reaction volume into several compartments and modelling diffusion between them.

3.4 Bayesian networks

A Bayesian network (BN) is a directed acyclic graph commonly used as a probabilistic modelling tool [Pearl, 1988]. Modelling chemical networks with BNs was introduced by [Sachs et al., 2002]. In a BN, variables (a molecular property) are represented as nodes in the graph. Directed edges express the dependence relation between nodes. A variable can be either discrete or continuous and may form a hypothesis, a known value (e.g., a concentration) obtained by experimental measurement or a latent variable. Variables which are not connected by edges are "conditionally independent".

If the state of a variable is known then the state of other variables can be predicted. This is accomplished through the use of:

$$p(x) = \sum {}_y p(x, y) \tag{3}$$

This formula sums the probabilities of all routes through the graph, thus allowing one to predict, with some probability distributions, the state of an unknown variable x. Continuous values for probabilities could be specified with a probability density function (e.g., [Needham et al., 2006] employs Gaussian distributions).

BNs have been used to reverse-engineer and infer the structure of biochemical networks [Sachs et al., 2002, Kim et al., 2003, Needham et al., 2006]. However, the setting of probabilities (learning) of BNs requires static experimental data, otherwise this may result in increasing the complexity of the task [Li and Lu, 2005, Chickering, 1996]. The solid foundation of BNs in statistics enables the handling of the stochastic behaviour of real chemical networks and noisy experimental measurements [de Jong, 2002]. Another attribute of using BNs is that they can be employed when incomplete or only steady-state data on the reaction network are available. In this common case, kinetic models have been found to be less useful [Woolf et al., 2005]. [Pe'er, 2005] discussed the various techniques to infer BN models from experimental data.

A computational analysis of a Bayesian network requires tracing through the nodes and edges. Its computational complexity grows linearly with the number of nodes.

3.5 Term rewriting systems

Regulated term rewriting is a basic principle of information processing. Biomolecules, their polymeric subunits or groups of similar biomolecules are interpreted as objects encoded by character strings (terms). Sets of term rewriting rules describe possible interactions among objects and system components (e.g., pathways or membrane structures). Each application of a rule performs a discrete step of a process. The terms as a whole contain all information about the system status. Term rewriting systems can run in a massively parallel manner considering nondeterministic recombinations. Classes of grammar systems, P-systems [Paun, 2002], broadcast language [Holland, 1975, Holland, 1992] and Alchemy based on the lambda calculus fall into this category [Fontana and Buss, 1994]. We demonstrate this modelling approach with the broadcast language (BL).

Holland originally proposed the BL formalism to assist his research on the "adaptive plan". Holland argued that the BL provides a straightforward representation for a variety of natural models such as biochemical networks.

The BL basic components are called *broadcast units* which are strings formed from the set of "monomers" $\Lambda = \{0, 1, *, :, \lozenge, \triangledown, \blacktriangledown, \triangle, p, '\}$. Molecular species are broadcast units which can be viewed as *condition/action* rules. Whenever a broadcast unit conditional statement (pattern matching expression) is satisfied, the computational action statement is executed, i.e., when an enzyme broadcast unit detects, in the environment, the presence of one or more specific substrate signal(s) then the broadcast unit broadcasts an output product signal. General signal processing can also be performed with broadcast units: e.g., a broadcast unit may detect a signal I and broadcast a signal I', so that I' is some modification of the signal I. The broadcast monomers/symbols encode for the pattern matching and computational/enzymatic functions of molecular species. In addition, broadcast symbols may act as both operators and operands addressing the reflexive nature of molecular species (i.e., a molecule may act as both an enzyme and/or substrate).

Limited stochastic elements are involved in the computational functions of broadcast units which result in a semi-stochastic behaviour at the system level. The modelling of genetic regulatory networks (which addressed only the regulatory/qualitative aspects of biochemical networks) using the BL was proposed by [Decraene et al., 2007]. Although possible, no quantitative studies have been previously reported to have been conducted with the BL. Finally the BL formalism does not account for spatial information.

The computational complexity for simulations depends on the functional structure of the molecular species. Here, complexity grows linearly with the term (string) length/complexity of the molecules.

3.6 Petri nets

Petri nets (PNs) are a graph-oriented formalism originally from formal software engineering. Developed in the early 1960s [Petri, 1962, Peterson, 1981], Petri nets provide a means to model and analyse systems, which comprise of properties such as concurrency and synchronisation. Petri nets consist of "places", "transitions", and "arcs". "Arcs" are used to connect the "transitions" and "places", "input arcs" connect "places" with "transitions", while "output arcs" start at a "transition" and end at a "place".

The modelling of biochemical networks with Petri nets was introduced by [Reddy et al., 1993]. Here, place nodes are used to represent molecular species (enzymes, compounds, ions etc.) and transition nodes to denote chemical reactions. Other elements can be defined to specify in detail the chemical reactions to occur [Pinney et al., 2003].

Ordinary Petri nets provide an accessible modelling tool with well-established analysis techniques. For this reason, the use of Petri nets for qualitative analysis of biochemical network is growing. However, due to their timeless nature, Petri nets are limited regarding dynamic network analysis.

A computational simulation of the dynamical Petri net behaviour takes simultaneously into account all places considering the number of (molecular) objects necessary to conduct a transition. The computational cost increases linearly with the number of places.

3.7 π-calculus

The π-calculus is a process calculus, which is a formal method for modelling concurrent communicating processes [Hoare, 1983, Milner, 1999]. The π-calculus provides a framework for the representation, simulation, analysis and verification of such systems. The π-calculus allows the application of algebraic reasoning in order to determine the equivalence between processes.

When modelling biochemical networks using π-calculus, molecules and their individual domains are treated as computational concurrent processes [Regev et al., 2001]. Complementary structural and chemical determinants correspond to communication channels. Chemical interactions and subsequent modifications coincide with communication and channel transmission.

The π-calculus provides a highly detailed description of network nodes. However, the basic π-calculus gives only a semi-quantitative view. A significant factor to be considered is the lack of an associated temporal dimension and as a result all interactions can occur with the same probability/rate. Extensions of the basic π-calculus address this limitation [Regev and Shapiro, 2004, Blossey et al., 2008].

The computational costs for simulations of Milner's pi-calculus heavily depend on the process structure. In the computationally worst case, a continuously forking or branching scheme, the reasoning requires exponential resources in the number of calculus primitives.

3.8 Agent-based models

In an agent-based model (ABM), several computational objects called *agents* are simulated to reproduce real phenomena within an artificial environment. ABMs originate from the late forties with the development of Cellular Automata [von Neumann, 1949] and have been extensively used in the following fields: complex systems, multi-agent systems, and evolutionary programming [Luck et al., 2004, Winikoff and Padgham, 2004]. An ABM is typically implemented with an object-oriented framework [Rumbaugh et al., 1991]. Each agent or class is defined with particular properties and methods. Agents are situated in space and time, interactions between with each other may occur following rules. Global and complex behaviour may emerge from these local agent-agent interactions and properties.

ABMs provide a flexible framework to: specify and refine with ease rules governing agent behaviours and interactions (e.g., using production rules or Boolean logic), secondly, to model emergent system or global behaviours [Ausk et al., 2006]. Preliminary works to model bio-chemical networks using ABMs appeared in the late nineties [Schwab and Pienta, 1997, Fisher et al., 1999]. ABMs consider the cell and its components as agents with cognitive capabilities. Two distinct ABM approaches are presented:

1. In Cellulat, which was developed by [Pérez et al., 2002, Gonzalez et al., 2003], a cell is seen as a collection of adaptive autonomous agents. Communication between agents is performed via propagating signals on a shared data structure, named "blackboard" referring to the blackboard architecture [Nii, 1986a, Nii, 1986b]. An agent receives a signal or a combination of signals from a designated blackboard level and transduces these into another signal (or set of signals) on the same or different blackboard level. Transduction mechanisms of the signal depend of the cognitive capabilities of the agent. A blackboard level could represent extracellular, membrane, cytosol or nucleus region, this enables the modelling of spatial organisation.

2. A second ABM is described where Learning Classifier Systems (LCS) are used to specify the agents' behaviour and interactions. LCS are systems constructed from condition-action rules called *classifiers*. LCS can be seen as a simplification of the broadcast language where classifiers are binary strings that can be viewed as IF/THEN statements. Holland's initial work was modified a number of times and at present many different varieties of learning classifier systems are available [Lanzi et al., 2002, Bull and Kovacs, 2005].

   LCS are commonly used as a machine learning technique. However [Holland, 2001] proposed an agent-based model where the agents' behaviour and adaptation are determined by the use of LCS. This work argued that LCS could be used to evolve a simple repertoire of condition-action rules to a more complex goal directed set of rules.

   In typical biochemical networks, interactions between molecules follow the same condition-action mechanisms. Thus Holland suggested that this approach could be used to model and simulate biochemical networks. His proposition to design chemical networks was to start with a LCS-based "over-general" model of a biological phenomenon (e.g., transformation of a healthy cell to a cancer cell). Then this general phenomenon could be refined through several iterations. At each iteration, the details (e.g., compartment level) of the occurring interactions can be specified. These iterations were continued until the desired network level/granularity was reached, where the submolecular objects are specified (e.g., protein ligand, receptor, ions etc.). This refining process highlights the top-down/hierarchical approach and descriptive power of LCS to model and simulate complex biochemical networks. Moreover this approach can be naturally coupled with Genetic Algorithms. This evolutionary feature may allow one to examine phylogenetic relationships between different reaction networks (where the signalling differences may be due to random molecular mutations). However no actual implementation and experimental examination of this system have ever been reported, therefore this proposal and associated potential benefits remain conjectural.

3.9 SBML & CellML

Modelling techniques may be employed in conjunction with a markup language to store generated models. The use of a standard format facilitates the analysis, visualisation, simulation and exchange of biochemical network models within the modelling community, providing opportunities for refinement and incorporation of new knowledge. So far, two approaches have emerged, resulting in the model-description languages SBML (Systems Biology Markup Language) [Hucka et al., 2004] and CellML [Lloyd et al., 2004], both based on the XML markup language [Bray et al., 2000].

– In SBML, a biochemical network is described in terms of the molecules taking part in it - termed species - and the reactions taking place between them. The present amount of each species can be expressed either in terms of its concentration or of the number of molecules present. Each reaction has an associated kinetic law, which defines the rate of the reaction depending on the present amount of its substrates. Additionally, the model can be subdivided into a fixed set of well-stirred compartments to include a non-hierarchical spatial component. Nevertheless SMBL models cannot specify fluxes between compartments at present (i.e., in SBML level 2 version 4 release 1).

– In CellML, a more general approach is taken, in which a model consists of components and connections between components. Each component can contain variables and a reaction between them, and connections are used to transfer the value of variables from one component to another.

Although CellML is following a slightly more general approach, it is not as widely used as SBML, for which a large collection of software tools is available (see `www.sbml.org` for a list of these tools). Additionally, the first model repositories have started to use SBML as a representation language, e.g., see the BIOMODELS database at `www.ebi.ac.uk/biomodels`. Therefore, SBML can be seen as the first emerging specification standard for biological models at the cellular level. Finally the use of such a common language provides the ability to analyse and complement intersecting information on differing compatible modelling techniques.

4 Comparison of approaches

In this section, we compare the previously introduced methods to model biochemical networks by using a set of defined criteria. Following this, a comparison table is presented to summarise this review. The intention is to determine a suitable modelling technique to be employed. We identify evaluation criteria with regards to stochasticity, time, granularity, space, topology and modularity.

4.1 Evaluation criteria

Relevant criteria are outlined here in order to compare the modelling techniques presented in Section 3:

- **Stochasticity**: This property reflects the range of possible processing scenarios that may be identified by the model.
  - *Deterministic*: The system behaviour purely depends on inherent data. No external or statistical fluctuation may occur and influence the system's dynamics. The system may only operate along one known path.
  - *Nondeterministic*: A number of alternative paths for system processing may exist which can be completely explored by the model. All possible scenarios are taken into account by the model in which no unanticipated events may affect the system's dynamics.
  - *Stochastic*: In contrast, stochastic models select one possible path in a random manner that can be based on a given probability distribution. This implies uncertainty (external and statistical fluctuation may be accounted for) and inhibits repeatability of systems runs.
- **Time**: This property describes how time is represented within the model.
  - *Atemporal*: When executed, the model remains static and introduces no temporal consideration.
  - *Events*: A sequence of pre-identified events defines the granularity of time. An event is an action within the system which characterises the progress of the system processing. Events are not necessarily equidistant in time. Dependencies between processes, their synchronization and concurrency may also be based on the interplay of events.

- *Discrete*: Temporal changes are characterised by fixed periodic intervals. A discrete time interval defines the smallest unit measuring the system's dynamic behaviour. Discrete time points allow one to express recursive formulation of the system processing. Discrete time may be referred as a global clock for the system.

- *Continuous*: Infinitesimal time intervals allow the finest granularity for measuring time represented by real numbers. Computer-based simulation techniques, by their nature, require an approximate discretisation of points in time.

– **Granularity**: This property designates how the molecules or particles are represented in the model. It refers to the abstraction level of their specification. The finer the granularity the more detailed the system that can be described. Granularity also constrains the level of monitoring capabilities.

- *Submolecular*: This level allows one to compose molecules by atomic specifiers or functional units (e.g., protein domains).

- *Molecular*: Molecules are considered as the smallest expressible unit. A mapping between the chemical substance and the assigned identifier (e.g., symbol) is either assumed or abstracted.

- *Species*: An enumerable amount of molecules having the same chemical substance is regarded as a species. This level of granularity enables one to quantify a molecular species as a whole within the system, however one cannot isolate an individual molecule of a given species.

- *Concentration*: Allows one to quantify the relative amount of a particular molecular species existing in a system. As represented by real numbers, transforming absolute molecular amounts into concentrations can require an approximation. Concentrations can be viewed as an approximation of the molecular species quantities.

– **Space**: When handling molecules of given granularity within a model, a system component which is analogous to a reactor is assumed. This component can provide space if the positioning of the molecules (within the reaction system) is taken into consideration.

- *Implicit*: Particle or molecule identifiers include spatial information, e.g., using an index. System components that control the evolution can be equipped with regulation schemes for updating this information. Here, a homogeneous distribution of the molecules within the reactor is assumed. In this "well-stirred" reactor, no boundaries are specified, and there is no explicit definition of space in the model.

  - *Compartmental*: A hierarchically nested or graph-based number of explicit compartments is distinguished. Each molecule is assigned to one of the specified compartments and can move from one compartment to another. Within each compartment, no further specification of molecular positioning is defined.
  - *Grid*: Apart from the compartmental structure, a spatial geometry is used to locate molecules more precisely. This way, discrete spatial distributions of molecules can be mapped using the model.
  - *Continuous*: The finest granularity of defining space is given by continuous values. Here, each molecule can be positioned arbitrarily within the reactor. Analogous to continuous time, computer-based simulations may require discretisation which would imply approximation.

– **Topology**: This designates the ability of the model to dynamically modify its structural components (e.g., pathway structure, dependencies between compartments, active membranes, receptor dynamics).
  - *Fixed*: A static system structure is assumed.
  - *Dynamic*: Principles or rules are defined that allow the system structure to change over time and space. These rules are a part of the model description.

– **Modularity**: This refers to the ability of the model to subdivide a given biological reaction system into functional sub-units (i.e., modules). The subdivision process is carried out through algorithmic strategies applied on the model. Modules are determined/classified according to specific properties (e.g., network topology/clusters, functions) across these sub-units. Modularity may facilitate the study of a system by examining sub-units independently instead of the system as a whole.
  - *No*: The whole reaction system is regarded as a monolithic entity which currently prevents the identification of sub-units.
  - *Hierarchical structure*: The sub-units are represented as nodes forming a tree-based structure. Modules communicate with each others (e.g., transmission of molecules from one sub-unit to another) via specified interfaces, typically through diffusion over transduction/communication channels.
  - *Graph-based structure*: These structures are a generalisation of tree-based structures which does not necessarily account for a hierarchical organisation.

4.2 Comparison table and discussion

As a summary of previous sections, a comparison table is presented (Table 1) which uses the criteria that were discussed above. The table provides an immediate comparison of differing modelling techniques and allows one to identify desirable attributes which may be necessary for modelling a specific biochemical system.

Model	Type	Time	Granularity	Space	Topology	Modularity	Complexity
Ordinary DE	deterministic	cont.	conc.	implicit	fixed	no	$O(N)$
Partial DE	deterministic	cont.	conc.	cont.	fixed	no	$O(N)$
Markov chains	stochastic	discrete	species	implicit	fixed	graph-b.	$O(c^N)$ †
Master equation	stochastic	cont.	species	implicit	fixed	graph-b.	$O(logN)$ ‡
Bayesian networks	probabilistic	atemporal	species	implicit	fixed	graph-b.	$O(N)$
Term Rewr. Syst.	algebraic	discrete	(sub)mol.	implicit	dynamic	graph-b.	$O(m)$
π-calculus	algebraic	events	molecular	implicit	fixed	graph-b.	$O(c^N)$ †
Petri nets	algebraic	events	molecular	implicit	fixed	graph-b.	$O(N)$
Cellulat	agent-based	discrete	(sub)mol.	compart.	dynamic	hierarch. or graph-based	$O(n)$
Agent-based LCS	agent-based	discrete	(sub)mol.	implicit	dynamic	hierarch.	$O(n)$

† Worst computational case. ‡ Using the Next Reaction Method, $O(n)$ otherwise.

Table 1: Comparison of modelling approaches with respect to previously defined classification scheme. Let N be the number of molecular species, n the number of molecular/object instances (i.e., compartments, molecules, sub-molecular components, etc.), m is the sum of molecular species (expressed as terms or strings) lengths and $c \geq 1$.

In modern systems biology, we notice an increasing refinement of available experimental data and resulting models. While early attempts to discover reaction network structures and properties typically focused on the steady-state analysis and probabilistic issues, current studies prefer to capture dynamical aspects through the identification of specific reactions or diffusion kinetics.

Nevertheless, the level of abstraction may significantly vary within the models available. Well-curated repositories tackle the challenge of integrating data and findings into assembled frameworks of established modelling techniques. We aim at providing a general classification of modelling techniques according to their capabilities and advantages from a user's point of view. The classification scheme should be as simple as possible and clearly state which kind of information can be obtained from a model and which cannot.

5 Bridges between approaches

Historically the development of each of the modelling techniques resulted in a number of different approaches being explored. A result of this was a difficulty to express each model in a "common language". A number of models have been developed which attempt to "bridge" between the principal modelling approaches:

- The *heterogeneous* approaches allow for a combination of two or more modelling approaches into a single model. These unified approaches combine the advantages of each individual modelling technique, and ultimately would allow a researcher to construct models addressing the individual needs. From Table 1 we presented the principal properties of the differing modelling techniques. Through heterogeneity it is possible to create models which have arbitrary combinations of these properties. This allows for more flexibility in model composition with regards to experimental constraints. These approaches facilitate the finding of intersections of described issues.

  For example, stochastic differential equations (SDEs) extend differential equations to express stochasticity through the introduction of a stochastic term $\xi_x(t)$ into the governing reaction equations. These terms are perceived as random perturbations to the deterministic system. Further examples of heterogeneous approaches include: the Stochastic π-calculus [Priami, 1995, Lecca and Priami, 2007] and the Metabolic P systems [Manca, 2007] which gives an example for embedding continuous kinetics into a multiset-based framework.

- Differing modelling techniques can be unified without the requirement to produce new heterogeneous modelling techniques. Although heterogeneous modelling may create interesting combinations of two or more modelling types, it still leaves us with the problem of developing yet another modelling type. This perpetuates the ongoing difficulty of interoperability across models, and may also lead to an increase in complexity of a given model by incorporating more information.

  An alternative approach is to transform existing models to embrace information interchange rather than creating more incompatible and independent modelling techniques. The simplest approach is to utilise a common language (e.g. SBML, CellML) which allows for efficient information storage and interchange and also provides the ability to analyse and complement intersecting information on differing compatible modelling techniques. Note that the SBML possesses a longer history than CellML and has subsequently become the standard language for storing biochemical networks models. Therefore employing the SBML as a means to migrate and disseminate biochemical network models is advocated.

6 Conclusion

Systems biology is focused on achieving detailed descriptions of intra- and intercellular processes. It aims at a comprehensive mathematical model along with simulation studies able to explain and predict biological functions as a whole. From today's perspective, much effort is still required to reach this objective. Although the amount of available biological data is rapidly growing, its analysis and integration into one consistent global framework presents a serious challenge. In this context, an appropriate parameterisation of model specifications coping with partially incomplete wetlab experimental results is needed. We contribute to overcome this insufficiency by presenting a more global view on modelling approaches, their similarities and differences. Along with the widely used deterministic and stochastic descriptions of the reaction network and its dynamical behaviour, we emphasise the growing impetus of algebraic and agent-based description techniques. From a comparison study of recently relevant model types within these major description techniques, we suggest that algebraic (including agent-based) frameworks provide most flexibility. Because of their discrete composition of structural entities, they can act at different levels of abstraction ranging from sub-molecular interactions up to summarised system global function. Embedding analytical or stochastic information is enabled either by heterogeneous models or by model transformation. We believe that stages of interoperability between models for biochemical processes might promote systems biology towards a unified approach for all facets of biological information processing.

Acknowledgements

This work was funded by the ESIGNET project (Evolving Cell Signalling Networks *in silico*, a European Integrated Project in the EU FP6 NEST initiative, contract no. 12789). Further funding from the German Research Foundation (DFG, grant DI852/4-2) is gratefully acknowledged.

References

[Alberts et al., 2003] Alberts, B. et al. (2003). *Essential Cell Biology.* Garland Publishing.

[Alon, 2007] Alon, U. (2007). *An Introduction to Systems Biology: Design Principles of Biological Circuits.* Chapman & Hall/CRC.

[Atkinson et al., 2009] Atkinson, K., Han, W., and Stewart, D. E. (2009). *Numerical Solution of Ordinary Differential Equations.* Wiley Series in Pure and Applied Mathematics. Wiley.

[Ausk et al., 2006] Ausk, B., Gross, T., and Srinivasan, S. (2006). An Agent-based Model for Real-Time Signaling induced in Osteocytic Networks by Mechanical Stimuli. *Journal of biomechanics*, 39(14):2638–2646.

[Bernstein, 2005] Bernstein, D. (2005). Simulating Mesoscopic Reaction-diffusion Systems using the Gillespie Algorithm. *Physical Review E*, 71(4):41103.

[Blossey et al., 2008] Blossey, R., Cardelli, L., and Phillips, A. (2008). Compositionality, Stochasticity, and Cooperativity in Dynamic Models of Gene Regulation. *HFSP Journal*, 2(1):17–28.

[Bray et al., 2000] Bray, T., Paoli, J., Sperberg-McQueen, C., Maler, E., and Yergeau, F. (2000). Extensible Markup Language (XML) 1.0. *W3C recommendation*, 6.

[Bull and Kovacs, 2005] Bull, L. and Kovacs, T. (2005). *Foundations of Learning Classifier Systems*. Springer.

[Chatterjee and Vlachos, 2005] Chatterjee, A. and Vlachos, D. G. (2005). Binomial Distribution Based Tau-leap Accelerated Stochastic Simulation. *Journal of Chemical Physics*, 122.

[Chavali et al., 2008] Chavali, A., Gianchandani, E., Tung, K., Lawrence, M., Peirce, S., and Papin, J. (2008). Characterizing Emergent Properties of Immunological Systems with Multi-cellular Rule-based Computational Modeling. *Trends in Immunology*, 29(12):589–599.

[Chickering, 1996] Chickering, D. M. (1996). Learning Bayesian Networks is NP-complete. In Fisher, D. and Lenz, H.-J., editors, *Learning from Data: AI and Statistics V*, chapter 12, pages 121–130. Springer.

[de Jong, 2002] de Jong, H. (2002). Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *Journal of Computational Biology*, 9(1):67–103.

[Decraene et al., 2007] Decraene, J., Mitchell, G. G., McMullin, B., and Kelly, C. (2007). The Holland Broadcast Language and the Modeling of Biochemical Networks. In *EuroGP*, pages 361–370.

[Dittrich et al., 2001] Dittrich, P., Ziegler, J., and Banzhaf, W. (2001). Artificial Chemistries - A Review. *Artificial Life*, 7(3):225–275.

[Emonet et al., 2005] Emonet, T., Macal, C., North, M., Wickersham, C., and Cluzel, P. (2005). AgentCell: a Digital Single-cell Assay for Bacterial Chemotaxis. *Bioinformatics*, 21(11):2714–2721.

[Eungdamrong and Iyengar, 2004] Eungdamrong, N. J. and Iyengar, R. (2004). Modeling Cell Signaling Networks. *Biol Cell*, 96(5):355–362.

[Fisher et al., 1999] Fisher, M. J., Paton, R. C., and Matsuno, K. (1999). Intracellular Signalling Proteins as Smart' Agents in Parallel Distributed Processes. *Biosystems*, 50(3):159–171.

[Fontana and Buss, 1994] Fontana, W. and Buss, L. (1994). The Arrival of the Fittest: Toward a Theory of Biological Organization. *Bulletin of Mathematical Biology*, 56(1):1–64.

[Fontana and Buss, 1996] Fontana, W. and Buss, L. (1996). The Barrier of Objects: From Dynamical Systems to Bounded Organizations. In Casti, J. and A., K., editors, *Boundaries and Barriers*, pages 56–116. Redwood City, MA. Addison-Wesley.

[Fritz, 1982] Fritz, J. (1982). *Partial Differential Equations*. Number 1 in Applied Mathematical Sciences. Springer-Verlag, New York, NY, 4th edition.

[Genoud and Metraux, 1999] Genoud, T. and Metraux, J.-P. (1999). Crosstalk in Plant Cell Signaling: Structure and Function of the Genetic Network. *Trends in plant science*, 4(9):503–507.

[Gibson and Bruck, 2000] Gibson, M. and Bruck, J. (2000). Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *J. Phys. Chem. A*, 104:1876–1889.

[Gillespie, 1976] Gillespie, D. (1976). A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comput. Phys*, 22(4):403–434.

[Gillespie et al., 1977] Gillespie, D. et al. (1977). Exact Stochastic Simulation of Coupled Chemical Reactions. *The journal of physical chemistry*, 81(25):2340–2361.

[Gillespie, 2001] Gillespie, D. T. (2001). Approximate Accelerated Stochastic Simulation of Chemically Reacting Systems. *Journal of Chemical Physics*, 115(4):1716–1733.

[Gomez et al., 2001] Gomez, S., Lo, S., and Rzhetsky, A. (2001). Probabilistic Prediction of Unknown Metabolic and Signal-Transduction Networks. *Genetics*, 159(3):1291–1298.

[Gonzalez et al., 2003] Gonzalez, P., Cardenas, M., Camacho, D., Franyuti, A., Rosas, O., and Lagunez-Otero, J. (2003). Cellulat: an Agent-based Intracellular Signalling Model. *BioSystems*, 68(2-3):171–185.

[Goutsias, 2006] Goutsias, J. (2006). A Hidden Markov Model for Transcriptional Regulation in Single Cells. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 3(1).

[Heinrich and Schuster, 1996] Heinrich, R. and Schuster, S. (1996). *The Regulation of Cellular Systems*. Chapman & Hall.

[Hoare, 1983] Hoare, C. A. R. (1983). Communicating Sequential Processes. *Commun. ACM*, 26(1):100–106.

[Holland, 1975] Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor MI: University of Michigan Press, 1st edition.

[Holland, 1992] Holland, J. (1992). *Adaptation in Natural and Artificial Systems*. MIT Press Cambridge, MA, USA, 2nd edition. See also Holland(1975).

[Holland, 2001] Holland, J. (2001). Exploring the Evolution of Complexity in Signaling Networks. *Complexity*, 7(2):34–45.

[Huang and Ferrell, 1996] Huang, C. Y. and Ferrell, J. E. J. (1996). Ultrasensitivity in the Mitogen-activated Protein Kinase Cascade. *Proc Natl Acad Sci U S A*, 93(19):10078–10083.

[Hucka et al., 2004] Hucka, M., Finney, A., Bornstein, B., Keating, S., Shapiro, B., Matthews, J., Kovitz, B., Schilstra, M., Funahashi, A., Doyle, J., et al. (2004). Evolving a Lingua Franca and Associated Software Infrastructure for Computational Systems Biology: the Systems Biology Markup Language (SBML) project. *IEE Proceedings Systems Biology*, 1(1):41–53.

[Kim et al., 2003] Kim, S., Imoto, S., and Miyano, S. (2003). Inferring Gene Networks from Time Series Microarray Data using Dynamic Bayesian Networks. *Briefings in Bioinformatics*, 4(3):228–235.

[Krauss, 2003] Krauss, G. (2003). *Biochemistry of Signal Transduction and Regulation*. Wiley-VCH.

[Lanzi et al., 2002] Lanzi, P. L., Stolzmann, W., and Wilson, S. W., editors (2002). *Advances in Learning Classifier Systems, 4th International Workshop, IWLCS 2001, San Francisco, CA, USA, July 7-8, 2001, Revised Papers*, volume 2321 of *Lecture Notes in Computer Science*. Springer.

[Laubenbacher and Stigler, 2004] Laubenbacher, R. and Stigler, B. (2004). A Computational Algebra Approach to the Reverse Engineering of Gene Regulatory Networks. *Journal of Theoretical Biology*, 229(4):523–537.

[Le Novère and Shimizu, 2001] Le Novère, N. and Shimizu, T. S. (2001). StochSim: Modelling of Stochastic Biomolecular Processes. *Bioinformatics*, 17:575–576.

[Lecca and Priami, 2007] Lecca, P. and Priami, C. (2007). Cell Cycle Control in Eukaryotes: A BioSpi model. *Electronic Notes in Theoretical Computer Science*, 180(3):51–63.

[Li and Lu, 2005] Li, L. M. and Lu, H. H.-S. (2005). Explore Biological Pathways from Noisy Array Data by Directed Acyclic Boolean Networks. *Journal of Computational Biology*, 12(2):170–185.

[Li et al., 2006] Li, X., Rao, S., Jiang, W., Li, C., Xiao, Y., Guo, Z., Zhang, Q., Wang, L., Du, L., Li, J., et al. (2006). Discovery of Time-delayed Gene Regulatory Networks Based on Temporal Gene Expression Profiling. *BMC bioinformatics*, 7(1):7–26.

[Lloyd et al., 2004] Lloyd, C., Halstead, M., and Nielsen, P. (2004). CellML: its Future, Present and Past. *Progress in Biophysics and Molecular Biology*, 85(2-3):433–450.

[Luck et al., 2004] Luck, M., Ashri, R., and d'Inverno, M. (2004). *Agent-Based Software Development.* Artech House, Inc., Norwood, MA, USA.

[Manca, 2007] Manca, V. (2007). MP Systems Approaches to Biochemical Dynamics: Biological Rhythms and Oscillations. *Membrane Computing, 7th International Workshop, WMC*, pages 86–99.

[Meng, 2004] Meng, T. (2004). Modeling and Simulation of Biological Systems with Stochasticity. *In Silico Biology*, 4(3):293–309.

[Milner, 1999] Milner, R. (1999). *Communicating and mobile systems: the &pgr;-calculus.* Cambridge University Press, New York, NY, USA.

[Needham et al., 2006] Needham, C. J., Bradford, J. R., Bulpitt, A. J., and Westhead, D. R. (2006). Inference in Bayesian Networks. *Nature Biotechnology*, 24(1):51–53.

[Nii, 1986a] Nii, H. P. (1986a). Blackboard Systems, Part One: The Blackboard Model of Problem Solving and the Evolution of Blackboard Architectures. *AI Magazine*, 7(2):38–53.

[Nii, 1986b] Nii, H. P. (1986b). Blackboard Systems, Part Two: Blackboard Application Systems, Blackboard Systems from a Knowledge Engineering Perspective. *AI Magazine*, 7(3):82–106.

[Paun et al., 2006] Paun, A., Perez-Jimenez, M., and Romero-Campero, F. (2006). Modeling Signal Transduction using P systems. *Lecture Notes in Computer Science*, 4361:100.

[Paun, 2002] Paun, G. (2002). *Membrane Computing.* Springer.

[Pearl, 1988] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[Pe'er, 2005] Pe'er, D. (2005). Bayesian Network Analysis of Signaling Networks: a Primer. *Sci STKE*, 2005(281).

[Pérez et al., 2002] Pérez, P. P. G., Gershenson, C., Cárdenas-García, M., and Lagunez-Otero, J. (2002). Modelling Intracellular Signalling Networks using Behaviour-based Systems and the Blackboard Architecture. *CoRR*, cs.MA/0211029.

[Peterson, 1981] Peterson, J. L. (1981). *Petri Net Theory and the Modeling of Systems.* Prentice Hall PTR, Upper Saddle River, NJ, USA.

[Petri, 1962] Petri, C. A. (1962). *Template:Kommunikation mit Automaten.* PhD thesis, Bonn: Institut für Instrumentelle Mathematik, Schriften des IIM Nr. 2. Second Edition:, New York: Griffiss Air Force Base, Technical Report RADC-TR-65–377, Vol.1, 1966, Pages: Suppl. 1, English translation.

[Pinney et al., 2003] Pinney, J. W., Westhead, D. R., and McConkey, G. A. (2003). Petri Net Representations in Systems Biology. *Biochem Soc Trans*, 31(Pt 6):1513–1515.

[Polyanin and Zaitsev, 2003] Polyanin, A. and Zaitsev, V. F. (2003). *Handbook of exact solutions for ordinary differential equations. 2nd. ed.* Boca Raton, FL: CRC Press. xxvi, 787 p.

[Ponzoni et al., 2007] Ponzoni, I., Azuaje, F., Augusto, J., and Glass, D. (2007). Inferring Association Rules Between Genes using a Combinatorial Optimization Learning Process and Adaptive Regulation Thresholds. *IEEE/ACM Transactions on Computational Biology and Bioinformatics. In press, to appear in.*

[Priami, 1995] Priami, C. (1995). Stochastic pi-calculus. *Comput. J.*, 38(7):578–589.

[Reddy et al., 1993] Reddy, V., Mavrovouniotis, M., and Liebman, M. (1993). Template:Petri Net Representations in Metabolic Pathways. *Proc Int Conf Intell Syst Mol Biol*, 1:328–36.

[Regev and Shapiro, 2004] Regev, A. and Shapiro, E. (2004). The π-calculus as an Abstraction for Biomolecular Systems. *Modelling in Molecular Biology*, page 219.

[Regev et al., 2001] Regev, A., Silverman, W., and Shapiro, E. Y. (2001). Representation and Simulation of Biochemical Processes Using the π-Calculus Process Algebra. In *Pacific Symposium on Biocomputing*, pages 459–470.

[Rumbaugh et al., 1991] Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., and Lorensen, W. (1991). *Object-Oriented Modeling and Design.* Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

[Sachs et al., 2002] Sachs, K., Gifford, D., Jaakkola, T., Sorger, P., and Lauffenburger, D. A. (2002). Bayesian Network Approach to Cell Signaling Pathway Modeling. *Sci STKE*, 148(EG10).

[Schwab and Pienta, 1997] Schwab, E. D. and Pienta, K. J. (1997). Modeling Signal Transduction in Normal and Cancer Cells using Complex Adaptive Systems. *Med Hypotheses*, 48(2):111–123.

[Soinov et al., 2003] Soinov, L., Krestyaninova, M., and Brazma, A. (2003). Towards Reconstruction of Gene Networks from Expression Data by Supervised Learning. *Genome Biology*, 4(1):R6.

[Van Kampen, 2007] Van Kampen, N. (2007). *Stochastic Processes in Physics and Chemistry.* North-Holland.

[Vera et al., 2007] Vera, J., Balsa-Canto, E., Wellstead, P., Banga, J., and Wolkenhauer, O. (2007). Power-law Models of Signal Transduction Pathways. *Cellular signalling*, 19(7):1531–1541.

[von Neumann, 1949] von Neumann, J. (1949). Theory and Organization of Complicated Automata. *Burks (1966)*, pages 29–87.

[Winikoff and Padgham, 2004] Winikoff, M. and Padgham, L. (2004). *Developing Intelligent Agent Systems: A Practical Guide.* Halsted Press, New York, NY, USA.

[Woolf et al., 2005] Woolf, J., Wendy, P., Laurence, D., Daley, Q., and Lauffenburger, A. (2005). Bayesian Analysis of Signaling Networks Governing Embryonic Stem Cell Fate Decisions. *Bioinformatics*, 21(6):741–753.

[Zwillinger (Ed.), 1992] Zwillinger (Ed.), D. (1992). *Handbook of Differential Equations.* Academic Press, San Diego, CA, 2nd edition.