

Petri Nets for Modelling Metabolic Pathways: A Survey

Paolo Baldan¹, Nicoletta Cocco², Andrea Marin² and Marta Simeoni²

¹ Dipartimento di Matematica Pura e Applicata, Università di Padova
via Trieste 63, 35121 Padova, Italy
email: baldan@math.unipd.it

² Dipartimento di Informatica, Università Ca' Foscari di Venezia,
via Torino 155, 30172 Venezia Mestre, Italy
email: {cocco,marin,simeoni}@dsi.unive.it

Abstract. In the last fifteen years, several research efforts have been directed towards the representation and the analysis of metabolic pathways by using Petri nets. The goal of this paper is twofold. First, we discuss how the knowledge about metabolic pathways can be represented with Petri nets. We point out the main problems that arise in the construction of a Petri net model of a metabolic pathway and we outline some solutions proposed in the literature. Second, we present a comprehensive review of recent research on this topic, in order to assess the maturity of the field and the availability of a methodology for modelling a metabolic pathway by a corresponding Petri net.

1 Introduction

Molecular biology suffers of the gap between a huge amount of data stored in large databases, worldwide collected through observations and experiments, and the lack of satisfying explanations and theories able to give them a sound biological meaning. To fill the gap, Computational Systems Biology [73, 74] advocates the integration of experimental results with both computational techniques and modelling formalisms in order to get a deeper understanding of biological systems.

Various formalisms have been proposed for modelling and analysing bio-systems: ordinary differential equations (ODEs), process calculi, boolean networks, Bayesian networks, bipartite graphs, stochastic equations (see, e.g., [40, 122, 42, 80] for some surveys). It is not trivial to choose among discrete or continuous, deterministic or probabilistic modelling techniques. A difficulty resides in the need of having, at the same time, abstraction and ease of use, detailed and complete descriptions. Some additional problems are intrinsic to bio-modelling: heterogeneous representation of information, incomplete knowledge, noise and imprecision in the data. Data availability strongly influences the choice of the model: ODE-models are often the first choice when all kinetic data are known, but also stochastic models can be used. In the absence of kinetic data, the choice is obviously restricted to qualitative models. Still it is worth remarking that, in

any case, qualitative models, and the analysis methods they are equipped with, can provide valuable information which complements or facilitates a quantitative analysis.

A model always needs to be validated. After that, the model can be used for studying the behaviour of the biosystem and for experimenting with it. The analysis techniques available for the adopted formalism may be fundamental both in the validation phase and when studying the properties of the modelled biosystem.

In this paper we focus on metabolic pathways, which are complex networks of biochemical reactions describing flows of substances. Deterministic continuous models based on ODEs have been largely used for representing and analysing the kinetics of such networks. On the other hand, stochastic discrete models have been mainly used for simulation [123]. They assume that the timing of reactions is determined by a random variable (either continuous or discrete) and that compound concentrations change either by discrete numbers of molecules (discrete state space models) or by continuous values (fluid models), corresponding to single reaction events. ODEs model a metabolic pathway at a macroscopic scale, while a stochastic system models it at a finer level of granularity. As a consequence, a stochastic model, although generally hard to analyse, can reveal interesting individual behaviours neglected by ODEs, which instead capture, in a sense, the average case.

In some seminal papers Reddy et al. [100, 98, 99] and Hofestädt [67] propose Petri nets (PNs) for representing and analysing metabolic pathways. Since then a wide range of literature has grown on the topic. PNs are a well known formalism introduced in computer science for modelling concurrent systems. They have an intuitive graphical representation which may help the understanding of the modelled system, a sound theory and many applications both in computer science and in real life systems (see [96, 101, 90, 47] for surveys on PNs and their properties). A PN model can be decomposed in order to master the overall complexity and it enables a large number of different analyses. Just to mention a few, one can determine conflicting evolutions, reachable states, cycles, states of equilibrium, bottlenecks or accumulation points. Additionally, once a qualitative PN model has been devised, quantitative information can be added incrementally.

Although PNs have been employed also for signalling and regulatory networks (see, e.g., [55, 95, 84, 64, 103, 33, 57]), here we focus on metabolic pathways only, since modelling problems are different for different kinds of networks. PNs seem to be particularly natural for representing metabolic pathways, as there are many similarities between concepts in biochemical networks and in PNs. They both consist of collections of reactions which consume and produce resources and their graphical representations are similar. Such similarities allow for a fruitful integration between analysis techniques developed for biochemistry and for PNs. Abstraction and compositionality issues, as studied in PN theory, may help in mastering the complexity of metabolic networks, furthermore many tools are available for visualisation, analysis and simulation of PNs.

The goal of this work is twofold. Firstly, we illustrate how information about metabolic pathways can be modelled and analysed using PNs; we point out some relevant representation problems along with some solutions proposed in the literature. Secondly, we present a comprehensive review of recent research on modelling and analysing metabolic pathways with PNs in order to assess the maturity of the field. We consider in particular some aspects: the type of PNs, the case studies considered, the analyses performed, the tools adopted, the use of the main biological databases, the level of automated support for translating pathways information into corresponding PN models and for analysing such models.

Other reviews on modelling metabolic pathways through PNs are presented in [58, 94, 83, 32, 75]. In particular, in [58] the modelling and analysis capabilities of basic PNs and of their extensions are discussed, illustrating how the *glycolysis pathway* can be modelled with stochastic, coloured and hybrid PNs. In [94] PNs and some of their extensions are presented together with a classification of the analyses they enable and of the biological processes they can model. Three case studies are considered which highlight some typical features of biological systems. These are modelled with five PN tools, which are then compared with respect to their analysis capabilities. The paper [83] presents a survey on qualitative and quantitative modelling and analysis of biological pathways through various types of PNs. Some practical examples of modelling by means of Hybrid PN and their extensions are discussed, showing how they can be used to produce biological hypotheses by means of simulations. We also recall [32], which surveys the basics of PN theory and some possible applications to metabolic, genetic and regulatory networks. An interesting overview on PN and on the main analysis techniques for validating a PN model and experimenting with it, is presented in [75] and it is illustrated by examples. A special emphasis is given to the qualitative analysis of the model.

The paper is organised as follows. In Section 2 we shortly illustrate metabolic pathways and the main databases collecting related information. In Section 3 we present an overview of the basics of PNs. In Section 4 we discuss the representation of metabolic pathways through PNs. We describe the main problems concerning qualitative and quantitative modelling and we illustrate some solutions proposed in the literature. In Section 5 we present the main PN based approaches for modelling and analysing metabolic pathways in the literature and discuss the indications resulting from this survey. Brief concluding remarks follow in Section 6.

2 Metabolic Pathways

The life of an organism depends on its *metabolism*, the chemical system which generates the essential components - amino acids, sugars, lipids and nucleic acids - and the energy necessary to synthesise and use them. The *flow of mass and energy* is the essential purpose of the system and *homeostasis* its central property,

since the organism has to maintain a steady level for the important metabolites while facing external and internal stimuli.

Subsystems dealing with some specific function are called metabolic pathways. Hence a *metabolic pathway* is a set of interacting *chemical reactions* which modify a principal chemical. Each chemical reaction transforms some molecules (*reactants* or *substrates*) into others (*products*), and it is catalysed by one or more *enzymes*. Enzymes are not consumed in a reaction, even if they are necessary and used while the reaction takes place. The product of a reaction can be the substrate of subsequent ones.

Regulation is important in metabolic pathways. Usually it is obtained by a feedback inhibition or by a cycle where one of the products starts the reaction again. Anabolic (constructive) or catabolic (destructive) pathways can be separated by compartments or by using different enzymes. A metabolic pathway contains many steps, one is usually irreversible, the other steps need not to be irreversible and in many cases the pathway can go in the opposite direction depending on the needs of the organism. *Glycolysis* is a good example of this behaviour: it is a fundamental pathway which converts glucose into pyruvate and releases energy. As glucose enters a cell, it is phosphorylated by ATP to glucose 6-phosphate in a first irreversible step, thus glucose will not leave the cell. When there is an excess of energy, the reverse process, the *gluconeogenesis*, converts pyruvate into glucose: glucose 6-phosphate is produced and stored as glycogen or starch. Most steps in gluconeogenesis are the reverse of those found in glycolysis, but the three reactions of glycolysis producing most energy are replaced with more kinetically favorable reactions. This system allows glycolysis and gluconeogenesis to inhibit each other. This prevents the formation of a *futile cycle*, i.e. when two metabolic pathways running simultaneously in opposite directions have no overall effect except for dissipating energy.

A metabolic pathway is usually represented graphically as a network of chemical reactions. In such a concise representation enzymes are often omitted. In order to completely characterise a metabolic pathway, it is necessary to identify its components (namely the reactions, enzymes, reactants and products) and their relations. The quantitative relations between reactants and products in a balanced chemical reaction are represented through its *stoichiometry*. For example the well-known reaction in which water is formed from hydrogen and oxygen gas is described by the equation $2H_2 + O_2 \rightarrow 2H_2O$, where the coefficients show the relative amounts of each substance. Each amount can represent either the relative number of molecules, or the relative number of moles¹. Each reaction is characterised also by an associated rate, represented by a *rate equation*, which depends on the concentrations of the reactants. The rate equation depends also on a *reaction rate coefficient* (or *rate constant*) which includes all the other parameters (except for concentrations) affecting the rate of the reaction.

For a pathway with n reactions and m molecular species, stoichiometries may be represented in a more compact although less informative form by a

¹ By definition, a mole of any substance contains the same number of elementary particles as there are atoms in exactly 12 grams of the ^{12}C isotope of carbon.

stoichiometric matrix with n columns and m rows. An element of the matrix, a *stoichiometric coefficient* n_{ij} , represents the degree to which the i -th species participates in the j -th reaction or, more precisely, the variation of the amount of the i -th substance due to the j -th reaction. By convention, reactants have negative values and products positive ones. In elementary reactions such values are integers (whole molecules), while in composite reactions some values may be fractions. The coefficients of the enzymes are equal to zero, since they are taken and released in the reaction. For an introduction to metabolism and to chemical networks we refer to general texts such as [120, 23].

The study of a metabolic pathway requires to combine information from many sources: biochemistry, genomics, life process descriptions, network analysis and simulation. A challenge of computational Systems Biology is to represent metabolic pathways with formal models supporting analysis and simulation. Such models are meant to give a better understanding of the processes in a pathway by studying the interactions among the pathway components and how these interactions contribute to the function and behaviour of the whole system. Systems Biology advocates a method consisting of various steps:

- translate experimental and theoretical knowledge into a model;
- validate the model;
- derive from the validated model testable hypotheses about the system;
- experimentally verify them;
- use the newly acquired information to refine the model or the theory.

The first two steps are very complex, they require to build a model of a metabolic pathway from available knowledge and to validate it. Such knowledge is continuously increasing and rapidly changing, and often stored in databases.

2.1 Databases for metabolic pathways

In this section we briefly consider the main databases collecting knowledge on metabolic pathways.

The *KEGG PATHWAY* database [11] (KEGG stands for *Kyoto Encyclopedia of Genes and Genomes*) contains the main known metabolic, regulatory and genetic pathways for different species. It integrates genomic, chemical and systemic functional information [70]. At present it contains around 96000 pathways, generated from 339 reference pathways which are manually drawn, curated and continuously updated from published materials. Pathways are represented by maps with additional information connected to such maps. Such information cover reactions, proteins and genes, and may be stored also in other databases. KEGG can be queried through a language based on XML [9], called *KGML* (KEGG Markup Language) [10].

Another important repository is the *BioModels Database* related to the SBML.org site [18]. It allows biologists to store, search and retrieve published mathematical models of biological interest. These are annotated and linked to relevant data resources, such as publications, databases of compounds and

pathways, controlled vocabularies. At present it contains 231 curated and 198 non-curated models, 32014 species, 39293 reactions and around 16492 cross-references. The models are coded in *SBML* (Systems Biology Markup Language) [18], a language based on XML.

There are also other free access databases containing information on metabolic pathways:

- *MetaCyc* [12, 30], which is part of BioCyc Database Collection [2]. It describes more than 1400 metabolic pathways from more than 1800 different organisms and it is maintained by exploiting the scientific experimental literature.
- *Reactome* [17] which is an expert-authored, peer-reviewed knowledge base of core pathways and reactions in human biology, with inferred orthologous events in 22 non-human species.
- *TRANSPATH*, which is part of BIOBASE [20]. It provides information on signalling molecules, metabolic enzymes, second messengers, endogenous metabolites, miRNAs and the reactions they are involved in.
- *BioCarta* [1] hosts a set of dynamic graphical models which integrate emerging proteomic information from the scientific community. It contains around 80 metabolic pathways.

Relevant information can be found also in other databases, not specifically dedicated to metabolic pathways, such as *BRENDA* [5, 31] (BRaunschweig ENzyme DAtabase), which is the largest publicly available enzyme information system worldwide; *ENZYME* [8], which is another repository of information relative to enzymes; *DIP* [6, 104], *MINT* [13, 34] and *BIND* (Biomolecular Interaction Network Database) [4] which are catalogues of experimentally determined interactions between proteins.

2.2 General problems in representing metabolic information

In spite of all the available information, it is very difficult to collect the relevant knowledge and to translate it into a model of a metabolic pathway. The data concerning the chemical reactions are generally not precise, coherent and complete, and this is especially true for kinetic information. Published data come from different microorganisms or different strains and they are produced over several years. Besides the kinetic laws of reactions are seldom published, since major substrates, cofactors and effectors are usually studied separately. The kinetics of many processes is almost completely unknown and modelling assumptions are highly speculative. Hence simplified approaches are used [23, 122].

An example is the following *Michaelis-Menten scheme*

$$V = \frac{d[P]}{dt} = \frac{V_{max} \cdot [S]}{[S] + K_m}$$

which is commonly used to describe an enzymatic reaction



where E is the enzyme, S the substrate, P the product. In the scheme $[S]$ and $[P]$ denote the substrate and product concentrations, V_{max} is the maximum reaction rate and K_m the Michaelis constant. The parameters V_{max} and K_m characterise the interactions of the enzyme with its substrate. This kinetic model is valid when the concentration of the enzyme is much smaller than the concentration of the substrate and it is based on the assumption that enzyme and substrate are in fast equilibrium with their complex, that is the concentrations of intermediate complexes do not change (steady state approximation). When kinetic data are available, a more accurate description could be obtained by decomposing the enzymatic reaction and by using the *law of mass-action* for single-step reactions. An interesting discussion on the use of these two models and on their assumptions can be found in [28]. A uniform distribution of the amount of metabolites in the system is generally assumed, but in some situations this is not true. For example, when there are *diffusion mechanisms* or *membranes* which alter the transportation of metabolites. In such cases more appropriate kinetic models must be used [23].

A further problem to solve in modelling regards the partitioning of a large metabolic network into pathways. On the one hand partitioning is necessary to master the complexity of the network, on the other hand it is a quite difficult task since biologists and biochemists base their definitions of pathways on intuitive biological criteria. As a consequence, in the databases we may find models of the same pathway which can significantly differ, not only in numerical parameters, but also in their structure. To overcome this problem, consensus pathways may be defined and they are represented in some databases such as in KEGG [11].

In the latest years there have been many attempts to attack the problems concerning data availability and network partitioning by computational means. *Data mining and information retrieval* have been used to extract available metabolic information from the literature in the Internet and to integrate them with the information from major public databases, by using genome sequence data, annotations, organism similarities and statistical techniques (see, e.g., [26, 78, 27]). *Genetic algorithms* have been used for predicting both the network topology and the parameters of biochemical equations. They have been used also in the reverse engineering problem of inferring the topology and the parameters of a network from time series data on the concentration of the different reacting components (see, e.g., [77, 38]). *Optimisation techniques* have been used for the computation of steady-states, dynamic optimal growth behaviour and network modularisation (see, e.g., [88, 25]).

Analysis techniques are essential for model validation and various methods have been developed for metabolic pathways. We can distinguish *methods based on system theory*, such as *metabolic control analysis* and *flux balance analysis* (see, e.g., [65, 48, 119, 45]), and *structural methods* which do not use information on metabolite concentrations or fluxes, by assuming that fluxes and pool sizes are constant. The analysis of *extreme fluxes* and the analysis of *elementary flux modes* belong to this second class and are used to determine the possible behaviours of the pathway (see, e.g., [110, 105]). They are based on the stoi-

chometric relations which restrict the space of all possible flux patterns in a network to a rather small linear subspace. Such a subspace can be analysed in order to capture also the functional subunits of the network (see, e.g., [111]). A good survey on *metabolic pathways analysis* techniques can be found in [106]. Based on stoichiometric data and reaction descriptions, they have been used to identify and validate metabolic pathways [108, 109] without requiring kinetic information.

3 Petri nets

In this section we give an overview of Petri nets (PNs) (see, e.g., [96, 101, 90]), describing the basic model and some extensions used for the representation of metabolic pathways.

3.1 Basic Petri nets

A (finite marked) *Petri net* (PN) is a tuple $N = (P, T, W, M_0)$ where:

- P is the set of places, $P = \{p_1, \dots, p_n\}$;
- T is the set of transitions, $T = \{t_1, \dots, t_m\}$;
- $W : ((P \times T) \cup (T \times P)) \rightarrow \mathbb{N}$ is the weight function;²
when $W(x, y) = k$, with $k > 0$, the net includes an arc from x to y with weight k ;
- M_0 is an n -dimensional vector of non-negative integers which represents the *initial marking* of the net.

PNs admit a simple graphical representation, where places are drawn as circles, transitions as rectangles, the presence of an arc (p, t) or (t, p) between a transition t and a place p , is represented by a directed arc, labelled with the corresponding weight $W(p, t)$ or $W(t, p)$, respectively. When the weight is 1 the label is usually omitted. A *marking* (or *state*) is an n -dimensional vector of non-negative integers $M = (m_1, \dots, m_n)$ which represents the amount of *tokens* in the places of the net N . A marking M is graphically represented by inserting in each place p_i a corresponding number m_i of black circles representing *tokens*. We write $M \leq M'$ when $m_i \leq m'_i$ for $i \in \{1, \dots, n\}$. A simple example of PN is given in Fig. 1.

The *input bag* or *precondition* of the transition t is the n -dimensional vector of non-negative integers $t^- = (i_1, \dots, i_n)$, where $i_j = W(p_j, t)$ for any $j \in \{1, \dots, n\}$. Dually, the *output bag* or *post-condition* of the transition t is an n -dimensional vector of non-negative integers $t^+ = (o_1, \dots, o_n)$, where $o_j = W(t, p_j)$ for any $j \in \{1, \dots, n\}$. Intuitively, t^- indicates, for every place of the net, the number of tokens needed to enable transition t . The firing of t removes such tokens and generates new ones, as indicated by t^+ . We call input (output) places of a transition t those places, which correspond to the non-zero elements of t^- (t^+). For instance, in Fig. 1, we have $t_1^- = (1, 0, 0, 0, 0)$, while $t_1^+ = (0, 1, 1, 0, 0)$. Similarly, $t_5^- = (0, 0, 0, 2, 1)$ and $t_5^+ = (1, 0, 0, 0, 0)$.

² We denote by \mathbb{N} the set of natural numbers, $\mathbb{N} = \{0, 1, 2, \dots\}$.

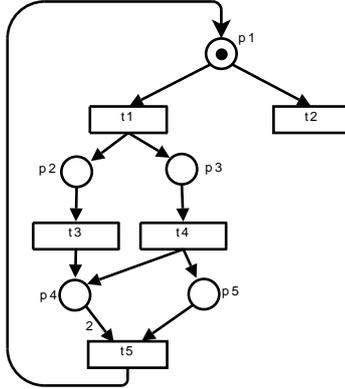


Fig. 1. A simple Petri net.

A *self-loop* of a transition t over a place p exists when $W(p, t) > 0$ and $W(t, p) > 0$, i.e., the transition deletes and produces tokens from the same place. A Petri net is called *pure* if it does not contain self-loops.

A transition t with input bag t^- is *enabled* by the marking $M = (m_1, \dots, m_n)$, if $t^- \leq M$. In this case t can *fire* and, as a consequence, the net marking changes from M to a new marking M' defined as follows:

$$M' = M - t^- + t^+,$$

and we write $M \xrightarrow{t} M'$.

Note that a marking can enable more than one transition. In this situation one of the enabled transitions is non-deterministically chosen and fired. It can also happen that some of the enabled transitions compete for a token; in this case they are in *conflict* and the execution of one of them can prevent the other ones from firing. For instance, consider the PN in Fig. 1. Its marking $M = (1, 0, 0, 0, 0)$ enables transitions t_1 and t_2 , which are in conflict. After firing e.g., transition t_1 , the new marking is $M' = (0, 1, 1, 0, 0)$ and t_2 is no longer enabled.

The *incidence matrix* of a PN N , denoted by \mathcal{A}_N , is the $n \times m$ matrix which has a row for each place and a column for each transition. The column associated with transition t is the vector $(t^+ - t^-)^T$, which represents the marking change due to the firing of t . In the PN literature the incidence matrix is sometimes defined as the transpose of the one considered here (see for example [90]).

The *reachability set* (or *state space*) of the net N is the set of all the markings of the net which are reachable by a firing sequence from the initial marking M_0 , and it is denoted by $R(N, M_0)$.

3.2 Behavioural properties of Petri nets

Representing a system as a PN allows for a formal analysis of its properties. We discuss now some fundamental properties which are of interest when PNs

model biological systems. First we focus on those properties which depend on the evolutions of the net from an initial marking. According to [90], these are referred to as *behavioural* or *marking dependent properties of a PN*. *Structural* or *marking independent* properties will be discussed in the next section. For the rest of this section N indicates a fixed PN, with m transitions, n places and an initial marking M_0 .

A first fundamental property of a system is *reachability*: given a marking M , determine whether $M \in R(N, M_0)$, i.e., check whether M is reachable via a firing sequence from the initial marking (reachability problem). Understanding whether the system will be able to reach some desired or undesired state is clearly of vital importance in most modelling application. A related problem is *coverability*: given a marking M , determine if there exists a reachable marking $M' \in R(N, M_0)$ such that M' covers M , i.e., $M \leq M'$. The problem of deciding whether a state is part of the reachability set of a net is known to be decidable [85], but of intractable complexity (it is easily proven to be NP-hard [47] and it is actually EXPSPACE-hard [79]). As a consequence of the intractability of the reachability problem, naive approaches to reachability analysis normally fail to be effective, a fact that motivates the interest and the huge literature on analysis methods for Petri nets.

Next fundamental property is *boundedness*. A net N is *bounded* if, during the evolution of the net, in each place the number of tokens will never exceed a fixed number k . More formally, N is called *k-bounded*, for some $k \in \mathbb{N}$, if for all $M = (m_1, \dots, m_n) \in R(N, M_0)$ and $i \in \{1, \dots, n\}$, $m_i \leq k$. We say that N is *bounded* if it is *k-bounded* for some $k \in \mathbb{N}$. When considering a PN model of a metabolic pathway, the validity of the conservation law should ensure that the net is bounded. However, often this property does not hold due to the presence of external metabolites for the considered pathway, which are usually assumed to be available in unbounded quantity. A more detailed discussion on the modelling of external metabolites can be found in Section 4.2.

The third fundamental property is *liveness*. A net N is *live* if, starting from any reachable marking, any transition in the net can be fired, possibly after some further firings. Formally, for all $M \in R(N, M_0)$ and $t \in T$, there exists $M' \in R(N, M)$ such that t is enabled at M' . A net N has a *deadlock* (or *dead state*) if it is possible to reach a marking in which no transition is enabled. The fact that a PN model of a metabolic pathway is live indicates that any reaction, at any moment may happen infinitely often, i.e., that the biological process will never restrict to a subprocess. A deadlock corresponds to the possibility for a metabolic process to become blocked, thus reaching a false equilibrium.

The *reachability graph* of a net N with initial marking M_0 , is a graph whose vertices are the elements of the reachability set $R(N, M_0)$ and whose arcs represent transitions, namely there is an arc from M to M' when $M \xrightarrow{t} M'$. The reachability graph is infinite when the number of reachable markings is infinite. In this case it can be useful to consider the *coverability graph*, which is a graph where nodes are labelled by markings, with the property that for each reachable marking $M \in R(N, M_0)$ there is a marking M' in the graph that

covers M . In order to keep the graph finite even for infinite-state Petri nets, the symbol ω is used to represent an unbounded number of tokens in a place. Conversely, for any marking M' in the coverability graph, there is a reachable marking $M \in R(M_0, N)$ such that $m_i = m'_i$, if $m'_i \neq \omega$, and $m_i > k$, for arbitrary k , if $m'_i = \omega$. Edges in the graph still represent the firing of transitions. The coverability graph is thus an abstraction of the reachability graph, where some information is lost. Hence its analysis normally provides bounds, rather than exact results.

Decision procedures for properties based on the reachability graph may be affected by the so-called *state space explosion* problem: a structurally small PN can have a huge reachability set due to the arbitrary interleaving of concurrent transitions. Indeed, for these nets the state space can grow faster than any primitive recursive function [118]. As a consequence, even if the reachability graph is finite, the analyses can become computationally unfeasible.

In order to tackle the state space explosion problem and to ease the analyses, many techniques have been proposed. Just to mention a few, reduction rules may be used to simplify a net while preserving its interesting properties (see, e.g., [90]). Binary decision diagrams, or generalisations of them, can be fruitfully combined with static analysis (e.g, [93]) to simplify the reachability analysis. Abstract interpretation may be used to reduce the size of the net (see, e.g., [49]) and partial order semantics to check reachability, coverability and absence of deadlocks and to model check behavioural logics for PNs [46].

3.3 Structural properties of Petri nets

The term *structural analysis of a net* refers to the analysis of those properties, called *structural properties*, which only depend on the net topology. Structural properties are independent of the initial marking and they can be often characterised in terms of the incidence matrix.

A common kind of structural analysis based on the incidence matrix aims to determine the so-called *invariants* of the net.

Let N be a PN, with m transitions and n places. A *T-invariant* (transition invariant) of N is an m -dimensional vector in which each component represents the number of times that a transition should fire to take the net from a state M back to M itself. It can be obtained as a solution of the following equation:

$$\mathcal{A}_N \cdot X = 0, \quad \text{where } X = (x_1, \dots, x_m)^T \quad \text{and } x_i \in \mathbb{N}, \text{ for } i \in \{1, \dots, m\}.$$

A T-invariant $X \neq 0$ indicates that the system can cycle on a state M enabling the cycle. The presence of T-invariants in a PN model of a metabolic network is biologically of great interest as it can reveal to the analyst the presence of steady states, in which concentrations of substances have reached a possibly dynamic equilibrium. As discussed in [52, 75], T-invariants admit two possible biological interpretations. On the one hand, the components of a T-invariant represent a multiset of transitions whose partially ordered execution reproduces a given initial marking. On the other hand, the components of a T-invariant may be

interpreted as the relative firing rates of transitions which occur permanently and concurrently, thus characterising a steady state.

The support of a T-invariant $X = (x_1, \dots, x_m)^T$ is the set of transitions corresponding to non-zero coefficients $\text{supp}(X) = \{t_i \mid x_i > 0\}$. The support of a T-invariant is *minimal* if there is no other T-invariant X' whose support is strictly included in that of X , i.e., such that $\text{supp}(X') \subset \text{supp}(X)$. For any minimal support there exists a unique T-invariant X with that support and such that for any other T-invariant X' , $X' \leq X$ implies $X = X'$. Such T-invariants are called *minimal support T-invariants* or simply *minimal T-invariants*. They form a basis for the set of T-invariants of the net, i.e., any T-invariant can be obtained as a linear combination of minimal T-invariants. In a PN model of a metabolic pathway, minimal T-invariants represent minimal sets of enzymes necessary for the network to function at steady state. They are particularly important in model validation techniques (see, e.g., [61, 75]) which will be discussed in Section 4.1.

A *P-invariant* (place invariant) or *S-invariant* (from ‘Stelle’, the German word for place) is a n -dimensional vector which can be obtained as a solution of the following equation:

$$\mathcal{A}_N^T \cdot Y = 0, \text{ where } Y = (y_1, \dots, y_n)^T, \text{ with } y_i \in \mathbb{N}, \text{ for } i \in \{1, \dots, n\}.$$

When a PN admits a P-invariant with all positive components, then the net is called *conservative* since the weighted sum of the tokens remains constant during the evolution of the net, i.e., for each marking of its reachability set. In a PN model of a metabolic network P-invariants correspond to the conservation law in chemistry. *Minimal (support) P-invariants* are defined similarly to minimal T-invariants. As discussed later, also minimal P-invariants are used as validators of a metabolic net model (see, e.g., [61, 75]).

A relevant role in the structural analysis of a PN is played also by *traps* and *siphons*. A *trap* S is a subset of places, $S \subseteq P$, such that transitions consuming tokens in S are a subset of those producing tokens in S , i.e., for any $t \in T$, $p \in S$, if $W(p, t) > 0$ then there exists $p' \in S$ such that $W(t, p') > 0$. As a consequence if a trap is initially marked, then it will never become empty.

Dually, a *siphon* S is a subset of places $S \subseteq P$ such that transitions producing tokens in S are a subset of those consuming tokens in S , i.e., for any $t \in T$, $p \in S$, if $W(t, p) > 0$ then there exists $p' \in S$ such that $W(p', t) > 0$. If a siphon is token free in some marking, then it will remain token free in all subsequent markings. In a biological model, traps may reveal the storage of substances which can occur, e.g., during the growth of an organism. Dually, siphons can characterise situations in which this storage, previously created, is consumed.

Structural properties can be useful also for the behavioural analysis of a net. For instance, a necessary condition for a marking M to be reachable is that the equation

$$\mathcal{A}_N \cdot X + M_0 = M \tag{1}$$

admits a non-negative integer solution.

Boundedness is obviously implied by *structural boundedness*, i.e., boundedness with respect to any possible initial marking. The latter can be characterised in terms of the existence of a strictly positive sub-P-invariant, namely an n -vector of strictly positive integers such that $\mathcal{A}_N^T \cdot Y \leq 0$.

For some subclasses of PNs, behavioural properties can be characterised in purely structural terms. For instance, condition (1) becomes also sufficient for reachability when dealing with acyclic nets, while for the so called free-choice PNs [41], liveness can be characterised in terms of traps and siphons.

A large number of tools have been developed for analysing many properties of PNs. A quite comprehensive list can be found at the Petri net World site [16].

3.4 Petri net extensions

The simplicity of the PN formalism, which is one of the reasons of its success, can also represent a limitation in the modelling of real systems. On the one hand, basic PNs, i.e. those described in the previous section, are not Turing powerful, on the other hand, they are not suited to model quantitative aspects of the behaviour of a system.

In order to overcome these limitations, several generalisations of the basic PN formalism have been proposed in the literature. Some extensions concern the qualitative aspects of the models, such as the addition of new kinds of arcs or coloured tokens, and they aim at increasing the expressive power or the modelling capabilities of the formalism. Other extensions introduce quantitative concepts, such as time and probability, allowing for the representation of temporal and stochastic aspects of biological systems, respectively.

In the following we review some extensions of PNs which have been used for the modelling of metabolic pathways.

– *PNs with test and inhibitor arcs.*

Test or *read arcs* allow a transition to verify the presence of a token in the source place without consuming it. *Inhibitor arcs*, instead, permit to model the fact that the presence of some tokens in a place inhibits the firing of a transition [21]. PNs with inhibitor arcs reach the expressive power of Turing Machines, and the reachability, liveness and boundedness problems become, in the general case, equivalent to the halting problem of the Turing Machine [71] and hence undecidable. The notions of P-invariant and T-invariant can be defined in the same way as for basic PNs. Under some restrictions, it is also possible to derive the coverability graph for this class of models [29].

In metabolic networks, test arcs may be used to model the role of the enzymes in reactions (see, e.g., [84,87]), since enzymes are needed for a reaction to take place, but not consumed. Inhibitor arcs may directly model the inhibitor function. This issue is further discussed in Section 4.1.

– *Self-Modifying or Functional Petri nets (FPNs).*

FPNs (see, e.g., [117]) give the possibility to define arc weights symbolically, as functions of the number of tokens in places. This allows arc weights to

evolve dynamically. When modelling metabolic networks, this feature can represent how variations in the concentration of substances may influence the kinetic rate of reactions [68].

– *Coloured Petri nets (CPNs)*.

CPNs [69] are a generalisation of basic PNs where the tokens are assigned a *colour* and the arcs can be labelled by conditions regarding these colours. Due to the possibility of making a distinction between tokens in the same place, CPNs can represent some system behaviours in a more compact and understandable way than basic PNs. In particular, the use of colours allows one to superpose different execution flows in the same net. As a consequence, a CPN can be structurally much smaller than a corresponding basic PN modelling the same process. For example, in [63, 121], CPNs are used to model a metabolic pathway where molecules of the same species are differentiated according to their origin and destination reactions.

– *Time Petri nets (TPNs)*.

In TPNs [86] a time interval $[a_t, b_t]$ is associated with each transition t . When t becomes enabled, it cannot fire before a_t time units have elapsed and it has to fire not later than b_t time units, unless t gets disabled by the firing of another transition. The firing of a transition t takes no time. TPNs are used to model or analyse time-related quantitative aspects of biological models, as in [62, 50, 97].

– *Stochastic Petri nets (SPNs)*.

SPNs [89] model time in a different way. Each transition is associated with a random variable which represents its firing delay. The expected value of the delay determines the execution rate of the transition, i.e., the number of firings per time unit when the transition is continuously enabled. If all the random variables are exponentially distributed (or more generally, if they take values in $(0, +\infty)$), then the reachability graph of the stochastic model is isomorphic to that of the untimed model. This makes some analysis tasks easier because standard techniques for basic PNs can still be used. Moreover, the reachability graph can be seen as a continuous time Markov chain whose states are the same as those of the reachability graph, and the transition from m' to m'' has a rate equal to the sum of the rates of the transitions that are enabled in m' and whose firings take the model to m'' [22, 24].

When a SPN models a metabolic pathway, the firing rate of a transition t may be a function of both the environment conditions (e.g., temperature, pressure, pH) and the concentrations of the reactants. SPNs are used for modelling biological systems, e.g., in [55, 81, 53, 60].

– *Continuous Petri nets (KPNs) and Hybrid Petri nets (HPNs)*.

KPNs are characterised by the fact that the state is no longer discrete. The integer numbers of tokens in places are replaced by non-negative real numbers, called marks. When modelling metabolic pathways, marks may represent the concentration of the molecular species. Continuous transitions have an associated firing rate, which expresses the “speed” of the transformation from input to output places. HPNs can represent both discrete and continuous quantities in the same model (see, e.g., [39]) since they have

both discrete and continuous places and transitions. Techniques for deriving a quantitative model of a biochemical network based on KPNs are proposed, e.g., in [52, 60, 28, 75].

Hybrid Functional Petri nets (HFPNs), defined in [84, 87], combine the features of HPNs and FPNs. Additionally, inhibitor and test arcs are used in order to ease the modelling of some biological functions.

Summing up, some PN extensions allow one to define models of biochemical reactions which capture also quantitative aspects. Some other extensions just lead to more compact and readable models than those which could be obtained with basic PNs. However, there is a price to pay. As a general rule, an increased modelling power makes the analysis tasks more difficult and, indeed, for extended formalisms there are fewer algorithmic analysis methods with an acceptable complexity or decent heuristics. For instance, when using CPNs, the discovery of T-invariants becomes more difficult. In [63, 121] it is shown how, in some situations, this problem can be faced by first decomposing the original net into sub-nets on the basis of the assigned colours, and then using standard algebraic techniques. When the number of colours is finite, CPNs can be unfolded into finite basic PNs. Hence several properties, like reachability, remain decidable. Nevertheless the unfolded model can be so large that the analysis is practically unfeasible.

Network simulation is generally necessary both for refining the model and for experimenting with it. KPNs, HPNs and SPNs can be used for simulation purposes. When modelling a metabolic network at a microscopic level, where a small number of molecules is involved and their individuality is important, SPNs are appropriate. As mentioned before, a SPN with an exponential time distribution induces a Markov chain and thus well-known analysis techniques are available. However, for complex systems these techniques can become practically unfeasible and simulation can be an interesting alternative. SPNs can be simulated by computing the delay for each enabled transition and by executing the transition with the smallest delay. This simulation corresponds to the Gillespie algorithm [54, 123], which is generally used in stochastic simulation of chemical reactions. The algorithm is simple but computationally expensive when the number of reactions/molecules increases. For this reason several modifications and adaptations have been proposed (see, e.g., [51]).

When modelling at a macroscopic level, the enormous number of molecules involved is better represented in a continuous way as a concentration, and KPNs or HPNs are generally more appropriate. In this case, the rate functions associated with transitions may follow known kinetic equations, which have been studied, under simplifying assumptions, for chemical reaction networks, such as the Michaelis-Menten or the mass action equation.

4 From Metabolic Pathways to Petri nets

In this section we discuss how to pass from a metabolic pathway description to a corresponding PN representation. We point out some modelling problems and

outline some solutions proposed in the literature. We also discuss the relations with existing analysis techniques for metabolic pathways in Systems Biology.

4.1 Correspondence between pathways and Petri nets

In order to give a PN representation of a metabolic pathway, the first step consists in providing a structural description. This process is guided by the natural correspondence between PNs and biochemical networks and it is well illustrated in [75]. Firstly, places are associated with molecular species, such as metabolites, proteins or enzymes. Transitions correspond to chemical reactions: places in the precondition represent substrates or reactants; places in the postcondition represent reaction products. The incidence matrix of the PN coincides with the stoichiometric matrix of the system of chemical reactions and arc weights in the PN can be derived by the stoichiometric coefficients. As a classical example, which dates back to [90], reaction $2H + O \rightarrow H_2O$ will lead to a PN with three places, corresponding to the substances H , O and H_2O , and one transition, which consumes two tokens from place H and one from place O and produces one token in place H_2O . The number of tokens in each place indicates the amount of substance associated with that place. It may represent either the number of molecules expressed in moles or the level of concentration, suitably discretised by introducing a concept of concentration level [53].

Once we have a qualitative model, quantitative data can be added to refine the representation of the behaviour of the pathway. In particular, extended PNs may have an associated transition rate which depends on the kinetic law of the corresponding reaction.

The following table shows the correspondence between metabolic pathway elements and Petri net elements.

Pathway elements	Petri net elements
metabolites, enzymes, compounds	places
reactions, interactions	transitions
substrates, reagents	input places
reaction products	output places
stoichiometric coefficients	arc weights
metabolites, enzymes, compounds quantities	number of tokens on places
kinetic laws of reactions	transition rates

Qualitative aspects. As an example of qualitative modelling, consider the two reactions in Fig. 2, which appear in the *Glycolysis Pathway*, as given in KEGG database.

By clicking on enzyme 3.1.3.11 in the KEGG map, it is possible to see the reaction associated with the enzyme, namely:



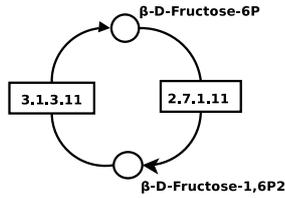


Fig. 2. Two biochemical reactions in the KEGG *Glycolysis Pathway*.

where the substrates are D-fructose 1,6-bisphosphate and water, and the products are D-fructose 6-phosphate and phosphate. Note that KEGG always uses the equal sign in reaction formulae even though the reaction is irreversible. The direction of a reaction is indicated by an arrow in the KEGG diagram.

The other reaction in the same figure, catalysed by the enzyme 2.7.1.11, is:



where the substrates are ATP and D-fructose 6-phosphate, and the products are ADP and D-fructose 1,6-bisphosphate. Also this reaction is irreversible.

If we represent each component of the substrate, each enzyme and each product of the reaction as a place of a Petri net, and chemical reactions by transitions, we obtain the PN of Fig. 3(a).

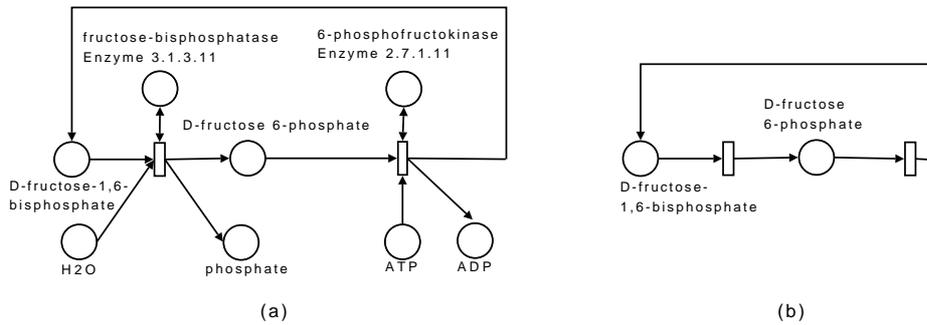


Fig. 3. PN associated with the biochemical reactions of Fig. 2.

Note that when enzymes are represented as normal substrates, the resulting PN is non-pure, i.e. it contains self-loops, since the enzyme is taken and released by the corresponding reaction. This is represented graphically in Fig. 3(a) by connecting the enzyme place to the transition corresponding to the associated reaction with a double arrow. Having a non-pure net may cause difficulties in the

analysis since self-loops are not represented in the incidence matrix of the net. In order to simplify the model and to avoid the introduction of self-loops, a typical modelling choice consists of omitting the explicit representation of enzymes. This is appropriate as long as their concentrations do not change as it happens, e.g., in signalling cascades. Alternatively enzymes can be treated as parameters, as suggested in [124].

Note that in the KEGG map of Fig. 2, substances such as H_2O , phosphate, ADP and ATP are not shown. They are ubiquitous molecules and their concentrations, which are assumed to be constant, are taken into account in the constants of the reaction rates.

By omitting places corresponding to enzymes and ubiquitous substances the net of Fig. 3(a) can be simplified, thus resulting in the PN of Fig. 3(b). In general, large and complex networks can be greatly simplified by avoiding an explicit representation of enzymes and by assuming that ubiquitous substances are in a constant amount. On the other hand, as an obvious drawback, processes involving these substances, such as the energy balance, are not modelled. For example a dephosphorylation, such as the decomposition of adenosine triphosphate (ATP) into adenosine diphosphate (ADP) and a free phosphate ion, releases energy, which is generally used by enzymes to drive other chemical reactions. This process cannot be observed in the simplified PN shown in Fig. 3(b).

The inhibitor function of some substances on a reaction can be directly modelled by using inhibitor arcs. However, inhibitor arcs generally make the properties of a PN much more difficult to check and the modeller might desire to avoid their use. If the net is bounded, one can get rid of inhibitor arcs by using a classical complementation technique [101]. Intuitively, a so-called complement place is introduced such that a token in such place represents the absence of an inhibitor biochemical substance. Inhibition of a process can also be modelled by using transitions which remove tokens needed by the process to take place, as discussed in [57]. Other inhibition patterns (negative feedback) in signalling networks and the corresponding PN encodings can be found, e.g., in [28]. Another approach consists in representing inhibition by quantitative information using the reaction rate. An example of this technique, applied to the modelling of negative feedback control, is illustrated later in this section, when discussing quantitative aspects.

Stoichiometry and invariants. In Systems Biology, many techniques for stoichiometric modelling, dynamic simulation and ODE-modelling of metabolic pathways have been developed. Petri nets allow for accommodating many of them into a single framework, while providing a natural graphical representation. Some clear correspondences exist between fundamental concepts used in modelling and analysing biochemical systems and their counterparts in PNs. In [111] they are discussed and summarised in a table which is partially reported below.

Biochemical systems	Petri nets
Conservation relations	P-invariants
Semi-positive (non-negative) conservation relations	Semi-positive P-invariants
Steady-state flux distributions	T-invariants
Elementary flux modes	Minimal T-invariants

The analysis of invariants in a Petri net model of a metabolic pathway has a fundamental role in checking its biological plausibility. Such validation techniques are well illustrated in [61, 75]. Moreover invariants may provide insights into the network behaviour.

A fundamental property to be verified is *CTI*, which requires that the PN model is covered by minimal T-invariants, i.e., that each transition belongs to the support of a minimal T-invariant. The fact that CTI holds means that every chemical reaction represented in the net occurs in some subprocess. The CTI property is a necessary condition for a bounded PN to be live. Besides each minimal T-invariant should be biologically meaningful and each biological behaviour should correspond to a T-invariant.

Similar validation criteria exist for minimal P-invariants. A net is said to satisfy *CPI* if it is covered by P-invariants, namely when every place is in the support of a minimal P-invariant. CPI can be expected to hold only for completely specified, closed systems. It will be violated whenever the system includes places corresponding to ubiquitous molecules and to external metabolites, as the components corresponding to these places will be null in any P-invariant. In fact, whenever a component y_p , corresponding to place p , is null, the P-invariant does not constrain the number of tokens in p (the representation of external metabolites is further discussed in Section 4.2). Nevertheless, each minimal P-invariant must have a biological interpretation and each compound conservation must correspond to a P-invariant.

The number of minimal T-invariants is generally high for large nets. As a consequence their biological interpretations becomes difficult to be grasped by inspection. Recently several techniques have been proposed for concisely representing the set of minimal T-invariants of a net and their relations. This may be quite helpful in easing the analyses. *Maximal Common Transition sets (MCT-sets)* have been introduced in [103]. They define a partition of the set of transitions T , where equivalence classes are formed by the transitions that participate in the same minimal T-invariants. More formally, two transitions t_i and t_j are in the same MCT-set if and only if, for every minimal T-invariant X , $t_i \in \text{supp}(X)$ if and only if $t_j \in \text{supp}(X)$. MCT-sets can be interpreted as the smallest biologically meaningful functional units of the net. Alternatively, minimal T-invariants can be classified by using standard clustering techniques. In this case a distance between T-invariants is defined by considering the similarity between their supports, computed, e.g., by comparing the number of common transitions with the total number of transitions of their supports. In [56] both MCT-sets and a well-known hierarchical clustering method (UPGMA) have been used for exploring T-invariants in model validation and in order to modularise the network.

A graphical representation of the minimal T-invariants in terms of a binary tree, the *Mauritius map*, has been proposed for the *knockout analysis* of a network. This is aimed at determining the most important transitions, namely those transitions whose knockout would destroy the largest number of functionalities (corresponding to minimal T-invariants) in the network. In a Mauritius map, starting from the root to the leaves of the tree the importance of transitions decreases. A detailed description of this technique can be found in [57].

Quantitative aspects. As observed in [66], kinetics is pivotal in a metabolic pathway and it should be represented in the corresponding PN model. Extended PNs can serve for this purpose. For instance in KPNs and HPNs the rates of chemical reactions can be very naturally modelled by associating marking dependent rates to continuous transitions. The underlying semantics can be expressed in terms of ODEs and well-know numerical techniques for their solution can be applied. Alternatively, stochastic discrete PNs can be used to obtain a finer grain model. In this case the exact analysis may become unfeasible for systems with very large state spaces and simulation becomes necessary. On the other hand, a SPN may point out biological behaviours that ODEs-based models are not able to catch.

Quantitative parameters can also be used to model control mechanisms in metabolic processes, such as positive or negative feedbacks. For instance, suppose that in a pathway the final product is an inhibitor of the enzyme that catalyses the first reaction (negative feedback control). We can define a state dependent rate for the PN transition corresponding to the first reaction of the pathway, such that its rate slows down when tokens accumulate in the place representing the final product. However, in general, defining appropriate transition rates is not easy, mainly because some relevant parameters, such as rate constants or kinetic laws, are not known or difficult to determine. This problem is not specific to PNs, but it rather applies to any formalism aiming at representing biological knowledge.

4.2 Problems in the PN representation of metabolic pathways

Although the correspondence between metabolic pathways and PN elements is rather straightforward, some problems arise in the construction of a PN representation of metabolic pathways. Below we review some of these problems and we present corresponding solutions in the literature.

P1: Modelling of spatial properties. In some cases, it is necessary to distinguish between compounds according to their location in the cell. For instance, in a cell, the ATP pools inside and outside the mitochondrion are different and their relative concentrations are determined through a selective transport process. This situation is modelled in [100, 98, 99] for the *adenine nucleotide transport system* by using two different places for ATP.

As discussed in [55, 32], spatial properties, such as positions, distances and compartments, are not naturally modelled with PNs. A system with various

compartments can be modelled by distinct interacting subsystems, where transitions represent the displacement of diffusing substances as proposed in [44, 72, 57]. As a drawback, the resulting model can become very large as it contains replicated information for the different compartments. Another solution, adopted in [63, 121], is to use colours to associate spatial information to substances.

P2: *Modelling of external metabolites.* In a metabolic pathway one can distinguish between internal and external metabolites. The former are entirely produced and consumed in the network, while the latter represent sources or sinks, that is, connection points with other pathways producing or consuming them. In a sense, external metabolites correspond to the network interface with the environment. The amount of external metabolites is usually assumed to be constant, thus implicitly assuming their unconditioned availability. They can be represented in several different ways.

A first possibility consists in simply not including the places corresponding to the external metabolites in the PN model. This can be fine for simulation purposes, while it can have a negative impact on the possibility of analysing the model, as, e.g., conservation laws can fail to hold.

When, instead, the model explicitly includes places associated with external metabolites, then they are characterised by the fact that all transitions in the net may either consume or produce tokens in such places. In the first case we talk of input metabolites and in the second case of output metabolites. They can be identified by those rows of the incidence matrix whose non-null coefficients have the same sign. In order to guarantee a correct behaviour for such places, various solutions have been proposed [63, 121, 124, 61, 76, 75], as illustrated below.

A first solution [61, 76, 75], consists in including, for any place in the net corresponding to an input metabolite a transition with empty precondition, generating tokens in such place. As transitions with empty preconditions can always fire, this corresponds to assume that input metabolites are always available. Similarly, outgoing transitions with empty postconditions can be added for output metabolites. Note that in this way no explicit assumption is made on the quantitative relations of input/output metabolites. Additionally, the obtained PN is unbounded, so it is not covered by P-invariants.

A different proposal, illustrated in [124], is to fill all the places corresponding to input metabolites with an infinite number of tokens, and to allow the places corresponding to output metabolites to accumulate an arbitrary number of tokens. In this way, situations of equilibrium which involve external metabolites will not be captured by T-invariants. Additionally, the net turns out to be unbounded and thus it is not covered by P-invariants.

Another possibility [124] consists in connecting places corresponding to output metabolites to those corresponding to input metabolites by additional transitions, in order to enable a circular flow. However, there is some arbitrariness in the choice of how places have to be connected.

A refined variant of the previous proposal, described in [63, 121, 61, 75], is to close the net by adding two auxiliary transitions, *generate* and *remove*,

connected by means of a place, *cycle*, as depicted in Fig. 4, which is taken from [61]. Transition *generate* produces tokens in the places modelling input metabolites, while transition *remove* consumes tokens from the places modelling output metabolites. The arc weights represent the stoichiometric relations of the sum equation of the whole network thus making explicit assumptions about the quantitative relations of input/output metabolites. The resulting net will be bounded and covered by P-invariants, when also the problem of adequate conflict resolution is addressed, as discussed in [63, 121, 75]. Note that, if quantitative aspects are included in the model, special attention must be devoted to the definition of the auxiliary transition rates, since they could impose unrealistic assumptions.

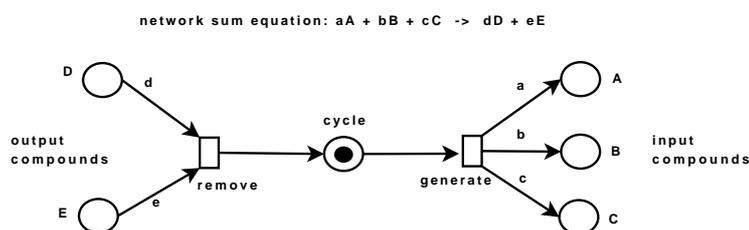


Fig. 4. Modelling external metabolites by adding the auxiliary transitions *remove* and *generate* and the auxiliary place *cycle*.

A further possibility, proposed in [124], is to create self-loops going from each output metabolite to the transition producing it, and from the transition related to each input metabolite back to the corresponding place. The resulting PN is no longer pure.

P3: Modelling of reversible reactions. A reversible reaction can occur in two directions, from the reactants to the products (forward reaction) or vice versa (backward reaction). The direction depends on the kind of reaction, on the concentration of the metabolites and on conditions such as temperature and pressure. The most important factor is *the activation energy* of the reaction, namely the minimal energy which is necessary for the reactants in order to start the reaction. Once started, the reaction proceeds. If the products have less energy than the reactants, the reaction produces energy which can activate other reactions. By the repeated activation of the forward reaction, the concentrations of reactants decrease and the concentrations of products increase. Correspondingly, the reaction rate decreases, since it is generally proportional to the concentrations of reactants, until eventually the backward reaction is activated. If the forward and backward reactions reach the same rate, then the reaction is in chemical (dynamic) equilibrium. Most of the reactions in a pathway are reversible. Each of them is normally viewed as a pair of distinct reactions, a forward one and a backward one,

which corresponds to a pair of transitions in the PN. If kinetic factors are not represented, the presence of the forward and backward transitions leads to a cyclic behaviour producing and destroying the same molecules, which might not be of biological interest. Still, this does not preclude the general suitability of qualitative analyses. In the T-invariant analysis, for example, minimal T-invariants corresponding to such cycles are usually considered as trivial T-invariants and ignored.

A solution for this problem, proposed in [63, 121], consists in adding specific biological knowledge to the PN by means of colours as discussed later in this section.

A more precise model is obtained if the kinetic factors can be represented. For example, if the forward reaction happens much faster than the backward one, the effect of the backward reaction may be unimportant. In [68, 43, 87] the authors face the problem by means of HPNs and HFPNs. Such models allow arc weights to be functions of the state of the network and thus the net behaviour can change depending on the reactant concentrations. As an example, in [68] a reaction is considered where for each element of the substrate we obtain two elements of the product; the reaction is catalysed by an enzyme and the concentration of the enzyme influences the reaction speed (see Fig. 5).

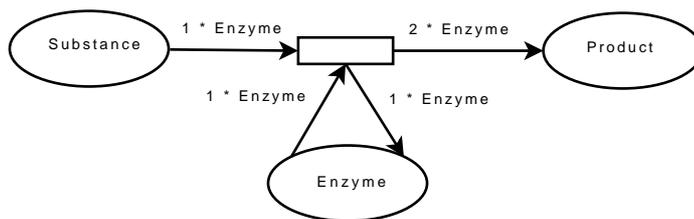


Fig. 5. FPN associated with a catalysed biochemical reaction, taken from [68]. In this model Substance, Product and Enzyme are variables which represent the number of tokens in the corresponding place. Arcs are labelled with functions of the variable Enzyme.

We conclude this section by remarking that, as shown by problem **P3**, when quantitative aspects of a PN can be considered, the resulting model is more accurate and sensible. However, often quantitative information are not available and, additionally, the simulation of a PN with quantitative information and the interpretation of the simulation results can be difficult tasks. For these reasons, the representation and the analysis is often restricted to qualitative aspects of the PN model and one may choose to add biological knowledge to the PN by using colours, as proposed in [63, 121].

As an example, suppose that the PN of Fig. 6(a) represents a fragment of a metabolic pathway and that the biological knowledge on such pathway ensures that either path $A \rightarrow B \rightarrow C$ or path $D \rightarrow B \rightarrow E$ will be followed each time. A SPN would represent this situation by using different rates for the transitions. If the active path is fixed, then the rates can be statically assigned. Otherwise, in the steady state assumption, marking dependent firing rates can be used to model the flow of tokens along the different paths. In the example, the rates of the transitions feeding C and E should be a monotone function of the number of tokens in A and D , respectively. A basic PN cannot represent this situation, since in place B it is not possible to discriminate whether tokens come either from A or from D and whether tokens proceed to C or to E .

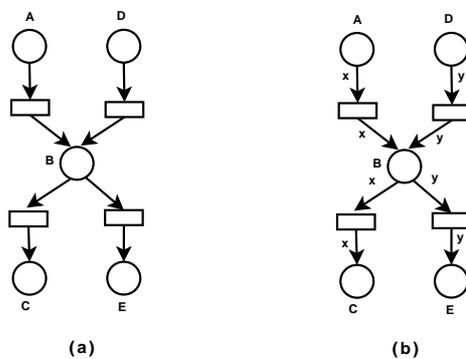


Fig. 6. Basic (a) and coloured (b) PNs representing a fragment of a metabolic pathway with the paths $A \rightarrow B \rightarrow C$ and $D \rightarrow B \rightarrow E$.

By using CPNs the colours x and y can be used to discriminate the two different paths, as shown in Fig. 6(b).

The colour assignment technique proposed in [63, 121] is based on the notion of *conflict place*, that is a place p with outgoing transitions which are in conflict. When a conflict place is identified in the basic PN, different colours are assigned to tokens belonging to different paths. Then, in order to separate different paths of the compound pathway and to distinguish among molecules in the same place according to their origin and destination, conditions are associated with reactions in a way that they will operate with tokens of some colours and ignore the presence of tokens of other colours. In [63, 121] the authors apply this technique to study the *glycolysis* pathway and show that the results of their analysis are more accurate than those obtained in [99] by using basic PNs.

5 Main Approaches in the Literature

In this section we present the main approaches in the literature to the modelling and analysis of metabolic pathways with PNs. The different proposals are grouped and discussed on the basis of the type of modelling and analysis performed, namely qualitative or quantitative. We conclude the section by pointing out the most relevant observations emerging from this literature review.

5.1 Qualitative modelling and analysis

The first approaches employing PNs for the modelling of biological networks have been presented in [100, 67, 99].

In [100] the authors consider the *fructose metabolism in liver* and present the corresponding basic PN. Petri net properties such as liveness, reachability, P- and T-invariants, are discussed along with their biological interpretation. In [99] the *combined metabolism of the glycolytic and the pentose phosphate pathways in erythrocytes* is considered. After presenting the corresponding basic PN, the authors provide a qualitative analysis in terms of P- and T-invariants, boundedness and liveness. Additionally, some reduction techniques are used for making the PN model smaller. This is done by ignoring details which are considered not essential for the analysis and by eliminating certain places and transitions without changing the overall behaviour of the model.

The focus of [67] is on the representation of metabolic processes with basic PNs, while the possibility of performing a qualitative analysis in terms of liveness, deadlocks and bottlenecks of the net is just mentioned. The paper considers the *Isoleucin pathway of E. coli* as a case study.

Basic PNs are also used in [124] for modelling the *Trypanosoma brucei glycolysis*. The authors illustrate the biological meaning of PN properties, such as siphons, traps, deadlocks and liveness and discuss the modelling problems **P2** and **P3** as presented in Section 4.2.

The use of CPNs for the design and qualitative analysis of metabolic pathways is proposed in [63, 121]. In particular, the authors consider the steady state analysis in the *combined glycolysis and pentose phosphate pathways in erythrocytes*, previously studied in [99], and discuss the role of T- and P- invariants. The analysis results turn out to be more accurate than those in [99] because, as discussed in Section 4.2, the use of colours allow the addition of specific biological knowledge to the PN model. In particular, a solution, already discussed in Section 4.2, is devised for the modelling problems **P1**, **P2** and **P3**. Some PN tools, such as Design/CPN [7] and the prototype tool SY (not publicly available), are used for the analysis.

In [92] an algorithmic methodology is proposed to study complex networks of chemical reactions without using the stoichiometry and the reaction rates. The methodology represents the network as a composition of CPNs, corresponding to molecular reactions and complex formations. Then the network is transformed into a digraph and, by further simplifications, into the list of all its minimal circuits. This operation allows for the identification of elementary biochemical

sub-networks and the study of the molecular flow; additionally it suggests new experiments which may be of interest. To demonstrate the suitability of the approach, the *Krebs cycle* is modelled and analysed.

An automatic translation of SBML models into PNs expressed in PNML (Petri Net Markup Language) [15] is proposed in [112]. The translation is supported by a prototype tool (not publicly available). The proposed methodology is illustrated on the *saccharomyces cerevisiae glycolysis* pathway taken from the BioModels database [3]. Using the Integrated Net Analyser (INA) [115], a qualitative analysis based on P- and T-invariants is performed on the resulting PN. The use of SBML models allows for the specification of both qualitative and quantitative aspects of biological systems, but the proposed translation does not include the quantitative ones, since, at present, PNML does not support extended PNs.

Three related papers are [61, 76, 75]. In particular [61], which exploits the techniques introduced in [63, 121], presents PN models for the *sucrose breakdown pathway in the potato tuber* and the *combined glycolysis and pentose phosphate pathways in erythrocytes*, and validate them by means of P- and T-invariants. The latter case study is modelled by CPNs. The paper discusses also problem **P2**, as presented in Section 4.2. The PN model of the *sucrose breakdown pathway in the potato tuber* is presented in more detail in [76] and validated by means of P- and T-invariants. The considered case study is of great research interest since it is not yet completely understood. In [75] the authors present the main qualitative PN analysis techniques and illustrate them on the *combined glycolysis and pentose phosphate pathways in erythrocytes*. The pathway is modelled using two PNs that differ for the modelling of the external metabolites (see problem **P2** in Section 4.2). The authors mainly focus on the analysis of qualitative properties, but also sketch the derivation of a KPN. They use public domain tools, namely Snoopy [19] and INA [115] for simulation and qualitative analysis, and IDD-CTL [116] and the Model Checking Kit [107] for model checking properties expressed in CTL.

A systematic approach for the modelling of regulated metabolic pathways with PNs is proposed in [114]. The approach is illustrated on the qualitative modelling of the *biosynthesis of tryptopham in E. coli*, taking into account two types of regulatory feedbacks. The idea is to separately build the basic PN models of the metabolic and regulatory pathways, which are then integrated into a unique structure. In particular, for the regulatory pathways, the authors propose to start with a specification in terms of a generic regulatory graph, followed by its parametrisation. Next the obtained model is mapped into a PN, called Multi-level Regulatory PNs. The complete methodology for regulatory pathways is illustrated in [33].

In [102] the authors consider a complex metabolic network describing the main part of the *human body iron homeostasis*, present the corresponding basic PN model and illustrate the results of an invariant analysis performed by using both MCT sets and cluster analysis. The case study concerns a biological process which is not yet fully understood. The analysis is helpful both for the validation

of the model and for the formulation of some biological hypotheses to be further studied.

5.2 Quantitative modelling and analysis

Various proposals of using extended PN models for the quantitative modelling and analysis of metabolic pathways can be found in the literature.

FPNs are used in [68] to represent some quantitative aspects of biological systems and to solve problem **P3** of reverse reactions, described in Section 4.2. The authors discuss the computational properties of the considered family of extended PNs and consider *glycolysis* as an example.

In [62] the authors employ TPNs to model and analyse a particular flux mode of the *combined glycolytic and pentose phosphate pathways in erythrocytes*. A reversible reaction is modelled by assigning suitable time intervals to the two corresponding transitions, so that the preferred direction fires more often than the other one. The authors apply the token game animation for a first validation of the model and then analyse P- and T-invariants and other structural and behavioural properties. The employed tools are PED [14] (now superseded by Snoopy [19]) and INA [115].

In [50] the authors propose a methodology for modelling metabolic pathways as TPNs for simulation purposes. They infer the metabolic network topology together with its kinetic parameters by querying the BRENDA database of chemical reactions, looking for all the sequences of reactions involving the initial (source) and final (sink) substances of interest. For the *glycolysis* case study such a query produces a very large set of networks. By applying some heuristic filtering rules, it is possible to obtain a single metabolic network, which is then translated into a TPN. The rate of each reaction is calculated with a kinetic function based on the Michaelis-Menten law and by using the kinetic parameters extracted from BRENDA. The resulting model, after a suitable reduction phase, can be used for simulation, performed with the tool Design/CPN [7].

HPNs are proposed in [37] for quantitatively modelling and simulating gene regulated metabolic networks. Gene regulation and metabolic reactions can be naturally represented by using both discrete and continuous places and transitions, while inhibitor arcs are used to represent inhibitors and to explore how enzymes are regulated. As a case study, the *urea cycle and its regulation* is considered, extracting the corresponding metabolic information from KEGG and BRENDA. A tool, *Visual Object Net++* (VON++) is employed for modelling and simulation of the HPNs. Other papers [35, 36] propose a framework for the integration of information extracted from different biology databases, aimed at easing the development and the execution of modelling and simulation tools. More specifically, the authors propose an environment to extract data from the main biology databases and automatically translate them into Petri net models. An extension of PNML [15] called BioPNML (Biologic PNML) is introduced in [36] for expressing biological data.

In [82, 84, 43, 87], the authors introduce HFPNs which extend basic PNs with features particularly suited for modelling biological pathways. HFPNs allow for

a natural representation of kinetic information since rate equations can be directly represented. Additionally, place values can be integers, reals, booleans, strings or vectors thus allowing for a natural representation of complex data. The authors apply HFPNs to genetic and signalling networks, exploiting the discrete elements to model control and regulation, and the continuous elements to model metabolic reaction rates. In particular, the *glycolytic pathway in E. Coli controlled by the lac operon gene regulatory mechanism* is modelled through HF-PNs. Besides the authors propose a software tool, called *Genomic Object Net* (GON), for representing and simulating biopathways modelled as HFPNs. The successor of GON, presented in [91], is called *Cell Illustrator* (CI). It is based on HFPNs and on a language specifically defined for the description of biological pathways, called CSML (Cell System Markup Language).

A methodology for deriving a quantitative model of a metabolic pathway in terms of TPNs is proposed in [97]. The authors introduce a technique to derive some kinetic parameters by assuming that the net is in a steady state and that the reaction rates are constant. Under these hypotheses and with the knowledge of the stoichiometric relations, one can fix the rate of an arbitrary chosen reaction and then derive the reaction rates of the whole PN by exploiting the T-invariants. Additionally, a technique is proposed to decide time dependent realisability of a transition sequence. However, these methods can lead to numerical problems; in this case approximated solutions must be adopted. The considered case study is the *sucrose breakdown pathway in the potato tuber*, taken from [76].

Other approaches in the literature focus mainly on the modelling of signalling or regulatory networks and are out of the scope of this literature review. However, some interesting proposals are worth to be mentioned for their innovative contribution in the research area of modelling biological pathways with PNs. In order to cope with the lack of information on concentrations and kinetic parameters, [113] presents a stochastic PN simulation tool called NASTY (not publicly available), equipped with a genetic algorithm for estimating unknown kinetic parameters. In [52, 28] the authors propose an approach which exploits both ODEs and PNs (basic and continuous). The papers [53, 60] propose a unifying framework for modelling and analysing biochemical networks with basic, stochastic and continuous PNs. In [59] the authors illustrate how to use invariant analysis in order to analyse data obtained by simulation with KPNs. A classification of T-invariants based on clustering techniques is introduced in [56] to cope with network complexity. The *T-clusters* thus obtained are used for modularising the network into biologically relevant functional units. The authors employ also the MCT-sets introduced in [103]. The papers [81, 72, 57] use different kinds of PNs for modelling purposes and different analysis techniques. However, they have a similar starting point and goal: by using literature and experimental data, they study biological pathways which are not yet fully understood. First, the available knowledge is modelled and the model is validated. Then the model is taken as a basis for a better understanding of the process under study and the analysis results are interpreted for giving new insights on system behaviour and for planning further experiments. In particular, in [57] the authors model

gene regulation of the Duchenne muscular dystrophy by using mainly their own experimental data, both for model development and for the design of new experiments. This is the paper in which Mauritius Maps, described in Section 4.1, are proposed and used for knockout analysis.

5.3 Summary and discussion

In this section we provide a synthesis of the approaches presented in Sections 5.1 and 5.2. Table 1 summarises their main characteristics. The table includes only the approaches for modelling metabolic pathways, while proposals concerning signalling or regulatory pathways mentioned at the end of Section 5.2 are omitted. The approaches appear in the table in the same order as they are discussed in Sections 5.1 (upper part of the table) and 5.2 (lower part of the table). Each row refers to a specific approach and the columns specify the following information:

1. References to the considered papers.
2. The types of Petri net adopted.
3. Pathways used as case studies. We highlight with a boldface font the pathways not yet fully understood or whose analysis required specific experimental knowledge.
4. The main source of information, that is, specific literature or databases such as KEGG or BRENDA.
5. Qualitative PN analysis planned or performed.
6. Quantitative PN analysis planned or performed.
7. Techniques for deducing or computing quantitative information, when needed.
8. Availability of automatic translations from metabolic pathways to PNs: do the approaches propose a semi-automatic translation? Is it supported by a tool? This column is strictly related to the fourth column, since automatising the translation requires the availability of a database as a source of information for metabolic pathways.
9. Tools adopted for qualitative or quantitative analysis, if explicitly mentioned.³

On the basis of the considered approaches and looking at Table 1 we can make the following observations.

- *There is no clear preference on the type of Petri net.* The choice of the model is clearly constrained by the data available for the pathway of interest. Qualitative models must be used when none or very few kinetic data are available. In this case, the precision of the analysis can be improved by conveying as much information as possible into the net structure. For example, the use of colours in CPNs makes it possible to get more precise results

³ Some papers do not mention explicitly the use of any specific tool, even if the complexity of the considered case studies clearly suggests that calculations have not been performed manually. Still, in this case the tool entry will be empty.

Paper	PN Type	Case study	Data Source	Qual. Analysis	Quant. Analysis	Quant. Inform. Integration	Translation Method. Tools	Tools
[100] [99]	basic	fructose metab. in liver, glycolysis and pentose phosphate	literature	T-/P-inv. boundedness liveness	no	-	no	-
[67]	basic	Isoleucin of E. coli	literature	no	no	-	no	-
[124]	basic	Trypanosoma brucei glycolysis	literature	T-/P-inv., traps, siphons	no	-	no	INA
[63] [121]	CPN	glycolysis and pentose phosphate	literature	T-/P-inv.	no	-	no	SY Design/CPN
[92]	CPN	Krebs cycle	literature	identification of minimal circuits	no	-	no	ad-hoc sw
[112]	basic	Saccharomyces cerevisiae Glycolysis	Bio-Models	T-/P-inv.	no	-	yes/prototype	INA
[61] [76] [75]	basic CPN	glycolysis and pentose phosp., sucrose breakdown in the potato tuber	literature	T-/P-inv. CTL	no	-	no	Snoopy, INA, model checkers
[114]	basic	biosynthesis of tryptopham in E.Coli and two regulatory feed-backs	literature	reachability	no	-	no	-
[102]	basic	human body iron homeostasis process	literature	T-/P-inv.	no	-	no	cluster analysis
[68]	FPN	glycolysis	literature	no	no	-	no	-
[62]	TPN	glycolysis and pentose phosphate	literature	T-/P-inv.	no	-	no	PED, INA
[50]	TPN	glycolysis	BRENDA	no	simulation	Michaelis Menten + BRENDA	yes/ no	Design/CPN
[37]	HPN	urea cycle and its regulation	KEGG BRENDA	no	simulation	literature	no	VON++
[82] [43] [84] [87]	HFPN	lac operon gene regul. mech. and glycolytic pathways in E. Coli	literature	no	simulation	Michaelis Menten + constants	no	GON
[97]	TPN	sucrose breakdown in the potato tuber	literature	T-inv.	time-dependent realisability	quant. data deduced by the T-inv.	no	-

Table 1. Summary of the approaches on metabolic pathways with PNs.

by representing additional biological knowledge without moving to a quantitative model (see, e.g., [63, 121]). Quantitative models, like HPNs, HFPNs, KPNs or SPNs, may be appropriate when information about the kinetics is available.

Then the properties of a pathway that one wish to analyse play a fundamental role in determining the choice. For instance, a transient analysis aimed at predicting the concentration of a substance clearly requires a quantitative model. Besides, it is worth noting that even if we are interested in a quantitative analysis and kinetic data are available, qualitative models can provide valuable preliminary or complementary information.

- *The considered case studies are mostly well known pathways, such as the glycolysis.* This might depend on the fact that most of the papers are aimed at introducing new PN based methodologies for modelling and studying metabolic pathways. Well known pathways are more suited for this purpose: on the one hand, the reader is likely to be familiar with the pathway and can focus on the peculiarities of the methodology proposed, on the other hand the availability of complete information on the pathway eases the validation phase. The recurrence of glycolysis and few other pathways may be due to the incompleteness of information, mainly on quantitative data, about other pathways.

There are some recent papers, like [61, 76, 97, 102] for metabolic pathways (in boldface in the table) and [81, 72, 57] for other kinds of pathways, which try to apply these techniques to original case studies, i.e., pathways which are not yet fully understood, and infer new biological insights from the PN analysis results.

- *Few approaches make use of databases as main references for metabolic pathways [50, 37, 112], while most approaches refer to specific literature.* The reason might be that databases contain partial information and some unreliable data. In the databases only a few well studied pathways are provided with all the necessary kinetic information. On the contrary papers describing specific experiments are generally more precise and complete.
- *Most of the approaches consider just one or a few specific case studies which serve as running examples.* Only the approaches [50, 112] propose a general translation methodology from metabolic pathways into PN models and [112] develops also a prototype tool, which is, however, not publicly available. This is related to the previous item: only the approaches which rely on biology databases can propose a translation methodology and some tool for its automation. Note that, in order to devise a semi-automatic translation from metabolic pathways to corresponding Petri nets, querying the databases is normally not sufficient and techniques for inferring missing information must be integrated into the framework.
- *The various approaches focus either on qualitative or on quantitative analysis, but not on both.* Qualitative analysis is usually preferred, since it is simpler and it does not require kinetic data. However, the integration of basic PN and KPNs is sketched in [75] and investigated more deeply in [52, 28]. Moreover, recent papers, like [53, 60] describe a complete modelling and

analysis framework taking advantage from the various techniques provided by different types of PNs (basic, stochastic and continuous).

- *The qualitative analysis, in most cases, is based on net invariants.* Invariant analysis has become a standard way to validate the PN model of a metabolic pathway. It can also help in identifying paths worth to be tested in experiments and it can provide new biological insights, as in [81, 102, 72, 57]. This is not surprising since the (minimal) T-invariant analysis in PNs corresponds to the (elementary) flux modes analysis which was already employed to validate metabolic pathways. A few approaches take into consideration other analysis techniques based on siphons, traps and liveness.
- *Quantitative analysis is often synonymous of simulation.* This might be caused by the complexity of PN models of metabolic pathways and the lack of kinetic information. The last issue can be faced by developing techniques for inferring the missing information. For example [97] derives quantitative information from the PN structure by assuming that the system is in a steady state.

We may further observe that when modelling complex metabolic processes, in general, it becomes necessary to represent also regulatory and genetic phenomena (see, e.g., [114]). When quantitative modelling is considered, this requires the integrated representation of both discrete steps and continuous behaviours. In these cases hybrid models, like HPNs or HFPPNs might be an appropriate choice (see, e.g., [37, 82, 84, 43, 87, 83]).

6 Conclusions

In this paper we have surveyed the literature on modelling and analysing metabolic pathways with PNs. This research area has been explored in search of a comprehensive methodology able to retrieve information on pathways from the main databases, to automatically translate them into corresponding PN models, and to employ PN tools for the qualitative and quantitative analysis of the nets thus obtained. First, we have described the natural correspondence between pathways and PNs, discussing some problems concerning the modelling and analysis of metabolic pathways with PNs, along with some solutions proposed in the literature. Then we have reviewed the literature on this research field, trying to evaluate the existing approaches with respect to a complete methodology in the sense described above. From this analysis, it emerged that a semi-automatic translation may be feasible when modelling only the structure of a metabolic pathway. In fact such data are generally available, their representation in PN form is rather natural and qualitative analysis techniques for model validation are well-established. On the contrary modelling quantitative information of a metabolic pathway poses serious problems. Data are largely incomplete and they have to be obtained from different sources or inferred with appropriate techniques. Even when such data are available, it is not easy to choose a PN extension able to represent those data and supported by suitable tools. In our opinion, a reasonable solution would be to define a modelling framework able

to adapt to different input sources and to different analysis requirements, as proposed in [35, 36]. Such framework should be able to integrate various existing approaches and tools and it would require a translation between the various input formats. In the future, format standardisation and a greater availability of quantitative data will improve the reliability of databases. We believe that this could make a semi-automatic translation from pathways to PNs really feasible.

Altogether, the natural representation of metabolic information as PN models, where qualitative and quantitative information can be integrated, the well known analysis techniques inherited by this formalism, the similarities between some biological and PN analysis methods, definitively suggest the appropriateness of PNs as a formalism for representing, validating and analysing metabolic pathways. Admittedly, while being quite promising, this seems to be still at a preliminary stage with respect to a comprehensive and automated methodology. The poor knowledge of the kinetic parameters plays an important role in this context. However such difficulties are inherent to biological systems and they affect any formalism and tool aiming at modelling bio-knowledge.

Acknowledgements. The authors are grateful to Marcantonio Bragadin of the Dipartimento di Scienze Ambientali - Università Ca' Foscari di Venezia, for enlightening discussions on metabolic pathways and biochemical issues. They are also greatly indebted to the anonymous referees for their extensive and valuable comments and suggestions on preliminary versions of the paper.

References

1. BioCarta: Charting Pathways of Life. <http://www.biocarta.com>.
2. BioCyc: database collection. <http://BioCyc.org>.
3. Biomodels Database. <http://www.ebi.ac.uk/biomodels>.
4. Biomolecular Interaction Networks Database. <http://bond.unleashedinformatics.com>.
5. BRENDA: The Comprehensive Enzyme Information System. <http://www.brenda-enzymes.info>.
6. Database of Interacting Proteins. <http://dip.doe-mbi.ucla.edu>.
7. Design/CPN: Computer Tool for Coloured Petri Nets. <http://www.daimi.au.dk/designCPN>.
8. ENZYME: enzyme nomenclature database. <http://www.expasy.ch/enzyme>.
9. Extensible Markup Language. <http://www.w3.org/XML>.
10. Kegg Markup Language manual. <http://www.genome.ad.jp/kegg/docs/xml>.
11. KEGG pathway database - Kyoto University Bioinformatics Centre. <http://www.genome.jp/kegg/pathway.html>.
12. MetaCyc Encyclopedia of Metabolic Pathways. <http://metacyc.org>.
13. MINT: The Molecular INTeraction database. <http://mint.bio.uniroma2.it>.
14. PED - a hierarchical Petri net editor. <http://www-dssz.informatik.tu-cottbus.de/index.html?software/ped.html>.
15. Petri Net Markup Language. <http://www.pnml.org>.
16. Petri net tools. <http://www.informatik.uni-hamburg.de/TGI/PetriNets/tools>.

17. REACTOME a curated knowledgebase of biological pathways. <http://www.reactome.org>.
18. SBML: Systems Biology Markup Language. <http://sbml.org>.
19. SNOOPY: a software tool to design and animate hierarchical graphs. <http://www-dssz.informatik.tu-cottbus.de/index.html?software/snoopy.html>.
20. TRANSPATH: The Pathway Database. <http://www.biobase-international.com>.
21. T. Agerwala. A complete model for representing the coordination of asynchronous processes. Hopkins Computer Research Report 32, John Hopkins University, 1974.
22. M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets*. Wiley Series in Parallel Computing. John Wiley and Sons, 1995.
23. P. Atkin and J. de Paula. *Atkins' Physical Chemistry*. Oxford University Press, 2006.
24. G. Balbo. Introduction to Generalized Stochastic Petri Nets. In M. Bernardo and J. Hillston, editors, *Formal Methods for Performance Evaluation*, volume 4486 of *Lecture Notes in Computer Science*, pages 83–131. Springer, 2007.
25. J. E. Beasley and F. J. Planes. Recovering metabolic pathways via optimization. *Bioinformatics*, 23(1):92–98, 2007.
26. S. Borger, W. Liebermeister, and E. Klipp. Prediction of enzyme kinetic parameters based on statistical learning. *Genome Informatics Series*, 1(17):80–87, 2006.
27. S. Borger, J. Uhlendorf, A. Helbig, and W. Liebermeister. Integration of enzyme kinetic data from various sources. *In Silico Biology*, 7(S1 09), 2007.
28. R. Breitling, D. Gilbert, M. Heiner, and R. Orton. A structured approach for the engineering of biochemical network models, illustrated for signalling pathways. *Briefings in Bioinformatics*, 9(5):404–421, 2008.
29. N. Busi. Analysis issues in Petri nets with inhibitor arcs. *Theor. Comput. Sci.*, 275(1-2):127–177, 2002.
30. R. Caspi, H. Foerster, C.A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S. Y. Rhee, A. G. Shearer, C. Tissier, T. C. Walk, P. Zhang, and P. D. Karp. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, 36(Database issue: D623-D631), 2008.
31. A. Chang, Scheer M., Grote A., I. Schomburg, and D. Schomburg. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Research*, 37(Database issue: D588-D592), 2009.
32. C. Chaouiya. Petri net modelling of biological networks. *Briefings in Bioinformatics*, 8(4):210–219, 2007.
33. C. Chaouiya, E. Remy, and D. Thiéffry. Petri net modelling of biological regulatory networks. *Journal of Discrete Algorithms*, 6(2):165–177, 2008.
34. A. Chatraramontri, A. Ceol, L. Montecchi Palazzi, G. Nardelli, M. V. Schneider, L. Castagnoli, and G. Cesareni. MINT: the Molecular INTeraction database. *Nucleic Acids Research*, 35(Database issue: D572-D574), 2007.
35. M. Chen. Modelling and Simulation of Metabolic Networks: Petri Nets Approach and Perspective. In *Proceedings of the European Simulation Multiconference on Modelling and Simulation*, pages 441–444, 2002.
36. M. Chen, A. Freier, J. Koehler, and A. Ruegg. The Biology Petri Net Markup Language. In *Promise2002*, pages 150–161. Lecture Notes in Informatics, 2002.
37. M. Chen and R. Hofestädt. Quantitative Petri net model of gene regulated metabolic networks in the cell. *In Silico Biology*, 3(0029), 2003.

38. E. J. Crampin, S. Schnell, and P. E. McSharry. Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. *Progress in Biophysics & Molecular Biology*, 86:77–112, 2004.
39. R. David and H. Alla. *Discrete, Continuous, and Hybrid Petri Nets*. Springer, 2005.
40. Hidde de Jong. Modelling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology*, 9(1):67–103, 2002.
41. J. Desel and J. Esparza. *Free Choice Petri Nets*. Cambridge University Press, 2005.
42. Y. Deville, D. Gilbert, J. van Helden, and S. J. Wodak. An overview of data models for the analysis of biochemical pathways. *Briefings in Bioinformatics*, 4(3):246–259, 2003.
43. A. Doi, S. Fujita, H. Matsuno, M. Nagasaki, and S. Miyano. Constructing biological pathway models with Hybrid Functional Petri Net. *In Silico Biology*, 4(0023):271–291, 2004.
44. A. Doi, M. Nagasaki, and H. Matsuno. Simulation-based validation of the p53 transcriptional activity with Hybrid Functional Petri Net. *In Silico Biology*, 6(0001), 2006.
45. J. S. Edwards, M. Covert, and B. O. Palsson. Metabolic modelling of microbes: the flux-balance approach. *Environmental Microbiology*, 4(3):133–140, 2002.
46. J. Esparza and K. Heljanko. *Unfoldings - A Partial order Approach to Model Checking*. EACTS Monographs in Theoretical Computer Science. Springer, 2008.
47. J. Esparza and M. Nielsen. Decidability issues for Petri Nets - a survey. *Journal Inform. Process. Cybernet. EIK*, 30(3):143–160, 1994.
48. D. A. Fell. Metabolic Control Analysis: a survey of its theoretical and experimental development. *Biochem. Journal*, 286:313–330, 1992.
49. P. Ganty, J.-F. Raskin, and L. Van Begin. From Many Places to Few: Automatic Abstraction Refinement for Petri Nets. In J. Kleijn and A. Yakovlev, editors, *Proceedings of ICATPN'07*, volume 4546 of *LNCS*, pages 124–143. Springer, 2007.
50. H. Genrich, R. Küeffner, and K. Voss. Executable Petri Net Models for the Analysis of Metabolic Pathways. *Proceedings of the Workshop on Practical Use of High-level Petri Nets*, pages 1–14, 2000.
51. M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry*, 25(104):1876–1889, 2000.
52. D. Gilbert and M. Heiner. From Petri Nets to Differential Equations - An Integrative Approach for Biochemical Networks Analysis. In *Petri Nets and Other Models of Concurrency - ICATPN 2006*, volume 4024 of *LNCS*, pages 181–200. Springer, 2006.
53. D. Gilbert, M. Heiner, and S. Lehrack. A Unifying Frameworks for Modelling and Analysing Biochemical Pathways Using Petri Nets. *Proceedings of the Workshop on Computational Methods in Systems Biology (CMSB)*, pages 200–216, 2007.
54. D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 25(81):2340–2361, 1977.
55. P. J. Goss and J. Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl. Acad. Sci. USA*, 95(12):6750–6755, 1998.
56. E. Grafahrend-Belau, F. Schreiber, M. Heiner, A. Sackmann, B. H. Junker, S. Grunwald, A. Speer, K. Winder, and I. Koch. Modularization of biochemical networks based on classification of Petri net t-invariants. *BMC Bioinformatics*, 9(90), 2008.

57. S. Grunwald, A. Speer, J. Ackermann, and I. Koch. Petri net modelling of gene regulation of the Duchenne muscular dystrophy. *BioSystems*, 92:189–205, 2008.
58. S. Hardy and P.N. Robillard. Modeling and simulation of molecular biology systems using Petri nets: modeling goals of various approaches. *Journal of Bioinformatics and Computational Biology*, 2(4):619–637, 2004.
59. S. Hardy and P.N. Robillard. Petri net-based method for the analysis of the dynamics of signal propagation in signaling pathways. *Bioinformatics*, 24(2):209–217, 2008.
60. M. Heiner, D. Gilbert, and R. Donaldson. Petri Nets for Systems and Synthetic Biology. In *Proc. of SFM'08*, volume 5016 of *LNCS*, pages 215–264. Springer, 2008.
61. M. Heiner and I. Koch. Petri Net Based Model Validation in Systems Biology. In *Petri Nets and Other Models of Concurrency - ICATPN 2004*, volume 3099 of *LNCS*, pages 216–237. Springer, 2004.
62. M. Heiner, I. Koch, and S. Schuster. Using time-dependent Petri nets for the analysis of metabolic networks. In R. Hofestadt, K. Lautenbach, and M. Lange, editors, *Workshop Modellierung und Simulation Metabolischer Netzwerke*, Preprint No.10, pages 15–21. Faculty of Computer Science, Otto-von-Guericke University of Magdeburg, 2000.
63. M. Heiner, I. Koch, and K. Voss. Analysis and Simulation of Steady States in Metabolic Pathways with Petri nets. *Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools (CPN'01)*, pages 15–34, 2001.
64. M. Heiner, I. Koch, and J. Will. Model validation of biological pathways using Petri nets - demonstrated for apoptosis. *Biosystems*, 75:15–28, 2004.
65. R. Heinrich and T.A. Rapoport. A Linear Steady-State Treatment of Enzymatic Chains. *Eur. Journal of Biochem.*, 42:89–95, 1974.
66. T. Hofer and R. Heinrich. A Second order Approach to Metabolic Control Analysis. *Journal of Theoretical Biology*, 164:85–102, 1993.
67. R. Hofestädt. A Petri net application of metabolic processes. *Journal of System Analysis, Modelling and Simulation*, 16:113–122, 1994.
68. R. Hofestädt and S. Thelen. Quantitative modeling of biochemical networks. In *Silico Biology*, 1(0006), 1998.
69. K. Jensen. *Coloured Petri Nets. Basic Concepts, Analysis Methods and Practical Use*. Monographs in Theoretical Computer Science. Springer, 1997.
70. M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi. KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, pages D480–D484, 2008.
71. K. Kant. *Introduction to Computer System Performance Evaluation*. McGraw-Hill inc., 1992.
72. J. Kielbassa, R. Bortfeldt, S. Schuster, and I. Koch. Modeling of the u1 snrnp assembly pathway in alternative splicing in human cell using petri nets. *Computational Biology and Chemistry*, 2008.
73. H. Kitano. Computational Systems Biology. *Nature*, 420:206–210, 2002.
74. H. Kitano. Systems Biology: a brief overview. *Science*, 295:1662–1664, 2002.
75. I. Koch and M. Heiner. Petri nets. In B. H. Junker and F. Schreiber, editors, *Analysis of Biological Networks*, Book Series in Bioinformatics, pages 139–179. Wiley & Sons, 2008.
76. I. Koch, B. H. Junker, and M. Heiner. Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. *Systems Biology*, 21(7):1219–1226, 2005.

77. J. R. Koza. *Handbook of Metaheuristics - Chapter 7: Automatic synthesis of topologies and numerical parameters*, volume 57 of *International Series in Operations Research and Management Science*. Springer New York, 2003.
78. W. Liebermeister, J. Uhlendorf, S. Borger, and E. Klipp. Automatic integration of kinetic data for metabolic network modelling. In *ICSB 2007*, pages 80–87. ACM, 2007.
79. R. J. Lipton. The reachability problem requires exponential space. 62, New Haven, Connecticut: Yale University, Department of Computer Science, Research, Jan, 1976.
80. J. Mandel, N.M. Palfreyman, J.A. Lopez, and W. Dubitzky. Representing bioinformatics causality. *Briefings in Bioinformatics*, 5(3):270–283, 2004.
81. W. Marwan, A. Sujatha, and C. Starostzik. Reconstructing the regulatory network controlling commitment and sporulation in *Physarum polycephalum* based on hierarchical Petri net modelling and simulation. *Journal of Theoretical Biology*, 236:349–365, 2005.
82. H. Matsuno, S. Fujita, A. Doi, M. Nagasaki, and S. Miyano. Towards Biopathway Modeling and Simulation. In *ICATPN 2003*, volume 2679 of *LNCS*, pages 3–22. Springer, 2003.
83. H. Matsuno, C. Li, and S. Miyano. Petri Net Based Descriptions for Systematic Understanding of Biological Pathways. *EICE Transaction on Fundamentals of Electronics, Communications and Computer Sciences*, E89-A(11):3166–3174, 2006.
84. H. Matsuno, Y. Tanaka, H. Aoshima, A. Doi, M. Matsui, and S. Miyano. Biopathway representation and simulation on hybrid functional Petri net. In *Silico Biology*, 3(0032), 2003.
85. E. W. Mayr. Persistence of vector replacement systems is decidable. *Acta Informatica*, 15:309–318, 1981.
86. P. M. Merlin and D. J. Farber. Recoverability of communication protocols—implications of a theoretical study. *IEEE Transactions on Communications*, 24(9):1036–1043, 1976.
87. S. Miyano and H. Matsuno. How to model and simulate biological pathways with Petri Nets - a new challenge for system biology. In *International Conference on Applications and Theory of Petri Nets, Bologna, Italy, 2004*.
88. C. G. Moles, P. Mendes, and R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Research*, 13:2467–2474, 2003.
89. M. K. Molloy. *On the Integration of Delay and Throughput Measures in Distributed Processing Models*. PhD thesis, UCLA, Los Angeles, CA, 1981.
90. T. Murata. Petri Nets: Properties, Analysis, and Applications. *Proceedings of IEEE*, 77(4):541–580, 1989.
91. M. Nagasaki, A. Doi, H. Matsuno, and S. Miyano. Petri Net Based Description and Modeling of Biological Pathways. In *Algebraic Biology 2005*, pages 19–31, 2005.
92. J.S. Oliveira, C. G. Bailey, J. B. Jones-Oliveira, D. A. Dixon, D. W. Gull, and M. L. Chandler. A Computational Model for the Identification of Biochemical Pathways in the Krebs Cycle. *Journal of Computational Biology*, 10(1):57–82, 2003.
93. E. Pastor, J. Cortadella, and M. Peña. Structural Methods to Improve the Symbolic Analysis of Petri Nets. In S. Donatelli and H.C.M. Kleijn, editors, *Proceedings of ICATPN'99*, volume 1639 of *LNCS*, pages 26–45. Springer, 1999.

94. M. Peleg, D. Rabin, and R. B. Altman. Using Petri net tools to study properties and dynamics of biological systems. *Journal of the American Medical Informatics Association*, 12:181–199, 2005.
95. M. Peleg, I. Yeh, and R. B. Altman. Modelling biological processes using workflow and Petri net models. *Bioinformatics*, 18(6):825–837, 2002.
96. J.L. Peterson. *Petri Net Theory and the Modelling of Systems*. Prentice-Hall, 1981.
97. L. Popova-Zeugmann, M. Heiner, and I. Koch. Timed Petri Nets for modelling and analysis of biochemical networks. *Fundamenta Informaticae*, 67:149–162, 2005.
98. V. N. Reddy. Modeling Biological Pathways: A Discrete Event Systems Approach. Master’s thesis, The University of Maryland, M.S. 94-4, 1994.
99. V. N. Reddy, M.N. Liebman, and M.L. Mavrovouniotis. Qualitative Analysis of Biochemical Reaction Systems. *Comput. Biol. Med.*, 26(1):9–24, 1996.
100. V. N. Reddy, M. L. Mavrovouniotis, and M. N. Liebman. Petri net representations in metabolic pathways. In *ISMB93: First Int. Conf. on Intelligent Systems for Molecular Biology*, pages 328–336. AAAI press, 1993.
101. W. Reisig. *Petri Nets: An Introduction*. EACTS Monographs on Theoretical Computer Science. Springer Verlag, 1985.
102. A. Sackmann, D. Formanowicz, P. Formanowicz, I. Koch, and J. Blazewicz. An analysis of the Petri net based model of the human bod iron homeostasis process. *Computational Biology and Chemistry*, 31:1–10, 2007.
103. A. Sackmann, M. Heiner, and I. Koch. Application of Petri net based analysis techniques to signal transduction pathways. *BMC Bioinformatics*, 7(482), 2006.
104. L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie, and D. Eisenberg. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research*, 32(Database issue: D449-D451), 2004.
105. C. H. Schilling, D. Letscherer, and B. O. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, 203:229–248, 2000.
106. C. H. Schilling, S. Schuster, B. O. Palsson, and R. Heinrich. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.*, 15:296–303, 1999.
107. C. Schroter, S. Schwoon, and J. Esparza. The Model Checking Kit. In *Proceedings of the 24th International Conference on Application and Theory of Petri Nets (ICATPN 03)*, volume 2697 of *LNCS*, pages 463–472. Springer, 2003.
108. S. Schuster, T. Dandekar, and D. A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnology*, 17(March):53–60, 1999.
109. S. Schuster, D. A. Fell, and T. Dandekar. A general definition of metabolic pathway useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18(March):326–332, 2000.
110. S. Schuster and C. Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 2:165–182, 1994.
111. S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinformatics*, 18(2):351–361, 2002.
112. O. Shaw, A. Koelmans, J. Steggles, and A. Wipat. Applying Petri Nets to Systems Biology using XML Technologies. In E. Kindler, editor, *Proc. of the Workshop on the Definition, Implementation and Application of a Standard Interchange Format for Petri Nets. Satellite event of ATPN*, pages 11–25, 2004.

113. O. Shaw, J. Steggles, and A. Wipat. Automatic Parameterisation of Stochastic Petri Net Models of Biological Networks. *Electronic Notes in Theoretical Computer Science*, 151(3):111–129, 2006. Proceedings of the Second International Workshop on the Practical Application of Stochastic Modeling (PASM 2005).
114. E. Simão, E. Remy, D. Thieffry, and C. Chaouiya. Qualitative modelling of regulated metabolic pathways: application to the tryptopham biosynthesis in E. Coli. *Systems Biology*, 21(2):ii190–ii196, 2005.
115. P.H. Starke and S. Roch. The Integrated Net Analyzer. *Humbolt University Berlin*, 1999. www.informatik.hu-berlin.de/starke/ina.html.
116. A. Tovchigrechko. *Model checking using interval decision diagrams*. PhD thesis, Dept. of Computer Science - BTU Cottbus, 2006.
117. R. Valk. Self-modifying nets, a natural extension of Petri nets. In G. Ausiello and C. Böhm, editors, *Proc. of ICALP'78*, volume 62 of *LNCS*, pages 464–476. Springer, 1978.
118. R. Valk and G. Vidal-Naquet. Petri nets and regular languages. *Journal of Computer and System Sciences*, 23(3):299–325, 1981.
119. A. Varma and B. O. Palsson. Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/Technology*, 12:994–998, 1994.
120. D. J. Voet and J. G. Voet. *Biochemistry*. Wiley, 2004.
121. K. Voss, M. Heiner, and I. Koch. Steady state analysis of metabolic pathways using Petri nets. *In Silico Biology*, 3(0031), 2003.
122. W. Wiekert. Modelling and simulation: tools for metabolic engineering. *Journal of Biotechnology*, 94(1):37–63, 2002.
123. D. J. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman and Hall/CRC, 2006.
124. I. Zevedei-Oancea and S. Schuster. Topological analysis of metabolic networks based on Petri net theory. *In Silico Biology*, 3(0029), 2003.