In Silico Biology 3 (2003) 367–387
IOS Press

Steady State Analysis of Metabolic Pathways Using Petri Nets

Klaus Voss[1], Monika Heiner[2] and Ina Koch[3,*]

[1]*Fraunhofer-Institute for Algorithms and Scientific Computing (SCAI) [GMD], D-53754 Sankt Augustin, Germany*
[2]*Computer Science Department, Brandenburg University of Technology at Cottbus, D-03013 Cottbus, Germany*
[3]*Department of Bioinformatics, Technical University of Applied Sciences Berlin, D-13347 Berlin, Germany*

Edited by E. Wingender; received 28 April 2003; revised and accepted 20 May 2003; published 13 September 2003

ABSTRACT : Computer assisted analysis and simulation of biochemical pathways can improve the understanding of the structure and the dynamics of cell processes considerably. The construction and *quantitative* analysis of kinetic models is often impeded by the lack of reliable data. However, as the topological structure of biochemical systems can be regarded to remain constant in time, a *qualitative* analysis of a pathway model was shown to be quite promising as it can render a lot of useful knowledge, e. g., about its structural invariants. The topic of this paper are pathways whose substances have reached a dynamic concentration equilibrium (steady state). It is argued that appreciated tools from biochemistry and also low-level Petri nets can yield only part of the desired results, whereas executable high-level net models lead to a number of valuable additional insights by combining symbolic analysis and simulation.

KEYWORDS: metabolic pathway, steady state, elementary mode, high-level Petri net, S-invariant, T-invariant

INTRODUCTION

With the rapidly growing amount of new experimental data, the modeling of biological pathways occuring in the cell regained great interest. For this challenge in biosciences, biologists need theoretical methods and computational tools in order to prove, analyse, compare, and simulate these complex networks for different organisms and tissues. The results are of major importance also for the biotechnology and the pharmaceutical industry.

*Corresponding author. Email: ina.koch@tfh-berlin.de.

Electronic publication can be found in *In Silico Biol.* **3**, 0031 <http://www.bioinfo.de/isb/2003/03/0031/>, 13 September 2003.

1386-6338/03/$8.00 © 2003 – IOS Press and Bioinformation Systems e.V. All rights reserved

"The main focus in the mathematical modeling in biochemistry has traditionally been on the construction of *kinetic* models. The aim of these models is to predict the system dynamics" [Heinrich and Schuster, 1998]. Their analysis is commonly based on the solution of systems of differential equations. In this way, numerous kinetic models for different metabolic systems and membrane transport processes have been developed (for a review, see Heinrich and Schuster, 1996). A severe restriction, often encountered in the construction of these models, is the imperfect knowledge of the kinetic parameters.

On the other hand, a *structural* analysis of metabolic pathways mainly deals with the topology of how substances are linked by reactions. A central role is played by stoichiometric matrices, which indicate how many molecules of each substance are consumed or produced in the single reactions. Their analysis is based on the solution of algebraic equations, and is independent of any kinetic parameter. Of particular interest are biochemical systems persisting in a *steady state* (see section "Steady state pathways, elementary modes"), i. e., in which the concentrations of their substances have reached an equilibrium. An *elementary mode* (this term has been coined in Schuster and Hilgetag, 1994) can be regarded as a minimal set of reactions (resp. of the enzymes catalyzing them) that can operate at steady state. Knowledge about the flux rates and the elementary modes of a system allows "to define and comprehensively describe all metabolic routes that are both stoichiometrically and thermodynamically feasible in a given group of enzymes" [Schuster *et al.*, 2000a].

A metabolic system can be modeled as a Petri net in a straightforward way, as has been demonstrated for low-level nets in Reddy *et al.*, 1993, and Hofestaedt, 1994, and for high-level nets in Genrich *et al.*, 2001. The Petri net structure then truly reflects the biochemical topology, and the incidence matrix of the net is identical to the stoichiometric matrix of the modeled metabolic system. Accordingly, the mentioned elementary modes correspond almost directly to the minimal T-invariants known from the Petri net theory. An actual account of the structural analysis of metabolic networks and the analogy to Petri nets is given in Schuster *et al.*, 2000b.

The use of Petri nets for modeling quantitative (kinetic) properties of biochemical networks, especially for genetic and cell communication processes, was discussed in Hofestaedt, 1994, and Hofestaedt and Thelen, 1998. Other contributions followed, using various types of Petri nets like stochastic nets [Goss and Peccoud, 1998; 1999] and hybrid nets [Matsuno *et al.*, 2000]. Executable high-level net models of metabolic pathways, and their (almost automated) construction, simulation, and quantitative analysis are described in Genrich *et al.*, 2001.

The application of Petri nets to this field began in the nineties with the publications of Reddy *et al.* [Reddy *et al.*, 1993; 1996]. They present a low-level (place/transition) net to model the structure of the combined glycolytic pathway (*GP*) and pentose phosphate pathway (*PPP*) of erythrocytes. They use the well-known algebraic methods to compute S- and T-invariants of the net. A thorough analysis of an extended form of this pathway was performed by Koch *et al.*, 2000, which forms the starting point for this paper.

For computing conservation relations (S-invariants) and elementary modes (T-invariants) of metabolic pathways, the software package METATOOL [Pfeiffer *et al.*, 1999] has been developed (by biochemists) and successfully applied in a number of cases. However, merely the integer weighted S-invariants are detected. Moreover, only the overall reaction equations, i. e., the net effects of a pathway execution, can be computed, and any consideration of its dynamics, in particular of the partial order of the reaction occurrences, is missing.

The main achievements reported in this paper rely on the use of executable high-level net models, an executable high-level net model, and on symbolic analysis. This allows to consider the following crucial aspects:

a. It is well known that the detection and interpretation of invariants can substantially improve the

understanding of systems. In the context of certain problems, however, the most interesting system properties are *not* invariant. In these cases, very often the divergence of these structures from an invariant is of major importance as it indicates a *defect* or *effect* of the substructure in question. We shall introduce and formally define these concepts for high-level Petri nets in section "Steady state pathways, elementary modes" and use them extensively, in section "Defects, effects and invariants", for the symbolic structural analysis of the quite complex sample pathway in section "Models of the glycolysis and pentose phosphate pathway".

b. In high-level nets, the model designer can distinguish tokens via their colors. This is a prerequisite for overcoming the restrictions of low-level nets or METATOOL, i. e., for both detecting hitherto unknown S-invariants of our model and determining its partial order dynamics, as shown in section "Defects, effects and invariants".

The software tool of our choice - for graphical editing, analyzing and executing the net models - is Design/CPN [Design].

STEADY STATE PATHWAYS, ELEMENTARY MODES

First some Petri net notions are recalled that we will apply to metabolic pathways later on. The algebraic analysis of Petri nets mainly relies on their invariants. However, in a great number of systems, like metabolic pathways, deviations from invariants deserve even more attention. Hence the following definitions [see Genrich, 2002] will turn out to be useful.

Let $\mathcal{N}$ be a colored (high-level) Petri net with the sets S of places and T of transitions. (A transition represents a whole class of transition occurrences, each one determined by a particular *binding* of the variables in the adjacent arc labels by color values). The *incidence matrix* C of $\mathcal{N}$ is an $|S| \times |T|$ matrix whose elements c_{ij} are the positive/negative labels of the arcs pointing from/to transition t_j to/from place s_i. An *S-* resp. *T-vector* is a vector with an entry for each $s \in S$ resp. $t \in T$. A *marking* of $\mathcal{N}$ assigns a multi-set of tokens (colors) to each place $s \in S$. At a marking M, usually several transitions are *enabled* to occur. One transition occurrence leads to a follower marking, enabling other transitions, and thus defines a *partial order* among them. The initial marking M^0 of $\mathcal{N}$ and all markings reachable from M^0 are called *states* of the net system.

The (symbolic) analysis of $\mathcal{N}$ is based on multiplying C with vectors of *transformations of expressions*. A *distribution* is a mapping transforming the elements of a color set D into linear combinations (with integer coefficients) of elements of a not necessarily different set D'. A *substitution* replaces variables by expressions in colors.

Let y be an S-vector such that, for every place S, the component y_s is a combination (list) of *distributions* of S, and all the y_s have the same range. Then the transpose matrix C^T can be multiplied by the vector (one column matrix) y. A product $c_{ij} \cdot y_{S_i}$ is the application of the distributions of s_i to the arc expression c_{ij} and yields an integer linear combination of tokens from the range color set.

Treating y as a one-column matrix, the product $C^T \cdot y$ is a T-vector whose entries are integer linear combinations of colors denoting the marking differences caused by the individual transitions. It is called the *defect* of y. A vector consisting only of zero elements $\mathbf{0}$ (= 0 * arbitrary color) is called *null vector* and denoted by O.

The S-vector y is an *S-invariant* of $\mathcal{N}$ iff $C^T \cdot y = O$.

An S-invariant represents a *state quantity* of the net system, i. e., a quantity which, starting from the initial state, is maintained during the whole life time of the system. It describes a *conservation rule*, as known from many areas in (natural) science. Such a mandatory S-invariant is a valuable means to detect

inconsistencies of a system specification or model.

A *process* π is a partially ordered set of transition occurrences leading from a state M_1 to a state M_2, $M_1 \xrightarrow{\pi} M_2$. Ignoring the order of occurrences yields a T-vector x of combinations of transition occurrences which is called the *action* performed by π. To be precise, every entry of such a T-vector is a combination of integer weighted *substitutions* of color variables by expressions in colors, and all variables have to be substituted by colors of the same set. Each substitution corresponds to a binding of the variables around a transition which determines the particular kind of its occurrence. A product $c_{ij} \cdot x_{t_j}$ consists of simultaneously applying the substitutions in x_{tj} to the variables in the arc expression c_{ij}, and yields an integer linear combination of tokens from the color set of s_i.

Treating x as a one-column matrix, the state difference $\Delta = M_1 - M_2 = C \cdot x$ is an S-vector whose entries are integer linear combinations of colors denoting the marking differences effected by x on the different places. It is called the *effect* of x.

The T-vector x is called a *T-invariant* of $\mathcal{N}$ iff $C \cdot x = O$.

A process $M_1 \xrightarrow{\pi} M_2$ performing a T-invariant leads from one state to the same state again ($M_1 = M_2$). It *re-generates* the state M_1, hence defines a *cyclic* process.

As mentioned in the introduction, the approach described in this paper concentrates on the mere structure of the pathways, i. e., on the topology of the interconnections of metabolites via enzymatic reactions. Hence, it is *structural* or *qualitative* as it does not deal with the kinetics of the reactions. Constructing a Petri net of such a pathway is straightforward, representing metabolites as places, reactions as transitions, and the stoichiometric relations by labelled directed arcs between them. Examples can be found in Reddy *et al.*, 1993, Hofestaedt, 1994, Reddy *et al.*, 1996, Koch *et al.*, 2000 (low-level nets), or in Genrich *et al.*, 2001, and section "Models of the glycolysis and pentose phosphate pathway" of this paper (high-level nets). In the following, such a net is called *the* Petri net model of the pathway. Figure 1 shows a sample high-level net model of a metabolic reaction. Jensen, 1992-1997, gives an excellent introduction into the theory and application of colored Petri nets.

In our metabolic pathways, a distinction is made between *external* and *internal* metabolites according to whether or not they are involved in reactions outside the system considered. External metabolites are called *sources* resp. *sinks* of the pathway if they are produced resp. consumed by those (external) reactions. A metabolic pathway is said to persist in a *steady state* if the concentrations of all internal substances

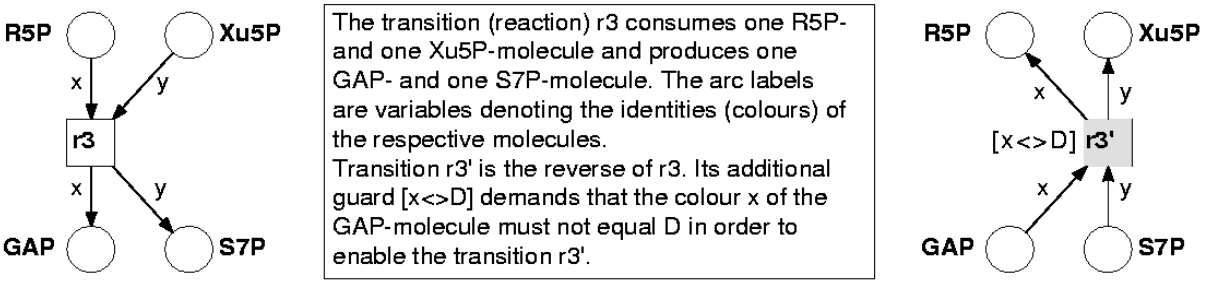

Figure 1: A sample reaction and its reverse. The transition (reaction) *r*3 consumes one R5P and one Xu5P molecule and produces one GAP and one S7P molecule. The arc labels are variables that denoting the identities (colors) of the respective molecules. Transition *r*3' is the reverse of *r*3. Its additional guard [*x*< >D] demands that the color x of the

GAP molecule must not equal D in order to enable the transition *r*3'.

have reached a dynamic equilibrium: for each internal metabolite, the total rate of its consumption equals that of its production. Assuming a constant activity of all enzymes involved in the system, many (but not all) metabolic pathways reach such a dynamic equilibrium after some time. That and how this happens, has been demonstrated for glycolysis, gluconeogenesis, citric acid cycle (TCA), and combinations of them in Genrich *et al.*, 2001, by simulation runs of quantitative high-level Petri net models. Structural analysis of metabolic systems in steady state aims at, among others, "elucidating relevant relationships among system variables" [Heinrich and Schuster, 1998] and does not rely on imperfectly known or doubtful kinetic data.

A formalization of steady state and related notions is given in Schuster *et al.*, 1996. For our paper, we need the following. The *stoichiometric matrix N* of a metabolic pathway with n metabolites and r reactions, is an $n \times r$ matrix where the element N_{ij} denotes the flux from the i-th metabolite to the j-th reaction, i. e., the amount $\delta c_i / \delta t$ of the metabolite concentration produced or consumed by that reaction. The stoichiometric matrix of a pathway precisely corresponds to the incidence matrix of its low-level net model.

A metabolic pathway is in *steady state* if and only if the reaction rates fulfill the condition

$\delta c_i / \delta t = \Sigma_{j=1}^{r} N_{ij} v_j = 0$, $i = 1, ...n,$ or, in matrix notation, $N \cdot v = O$,

for an integer vector $v = (v_1, ..., v_r)^T$, called *flux vector*, where a component v_j is an integer weight factor of the j-th reaction.

"Flux modes" constitute the core concept of the algebraic analysis of steady state metabolic pathways. "An *elementary flux mode* is a minimal set of enzymes that could operate at steady state, with the enzymes weighted by the relative flux they carry. 'Minimal' means that if only the enzymes belonging to this set were operating, complete inhibition of one of these enzymes would lead to cessation of any steady-flux in the system" [Schuster *et al.*, 2000a].

Before relating biochemical analysis methods to the corresponding Petri net algorithms, two particular questions have to be discussed. Firstly, in steady state analysis, those processes are of particular interest that start with the source substances of the investigated pathway and finish with its sink substances. For these external metabolites, constant concentrations have to be assumed to reach a steady state. Using METATOOL, this is achieved by excluding external metabolites from the stoichiometry matrix, but including the reactions affecting them: hence, their concentrations remain unchanged. In contrast, in our net models, we include the external substances and introduce an extra transition *StartEnd* that closes the pathway to a cycle by supplying the initially needed substrates and consuming the finally produced ones. This measure enables us to also identify those *internal* metabolites requiring initial markings and to compute their amounts.

Secondly, we have to take into account the reversibility of reactions. An obvious solution is to admit negative factors in the T-vector to denote the occurrences of reversible reaction transitions in the backward direction. This would lead to T-vectors $x \not\geq O$, not satisfying the standard definition of T-invariants for Petri nets. However, instead of deviating from this definition, we introduce, for every reversible transition t, an additional complementary (reverse) transition t' to the net (see Figure 1). Doing that, we get - for each reversible reaction t - a potentially endless loop (t, t', t, t', ...) which is biochemically meaningless. This slight disadvantage can be turned into an advantage in high-level nets where we can discriminate the directions of certain reactions according to the flux modes to which they belong. Thus we can model the situation in the cell, where the direction of a reaction depends on the need of the cell controlled by the metabolite concentrations.

MODELS OF THE GLYCOLYSIS AND PENTOSE PHOSPHATE PATHWAY

The *glycolysis pathway* (*GP*) is a sequence of reactions that converts glucose into pyruvate with the

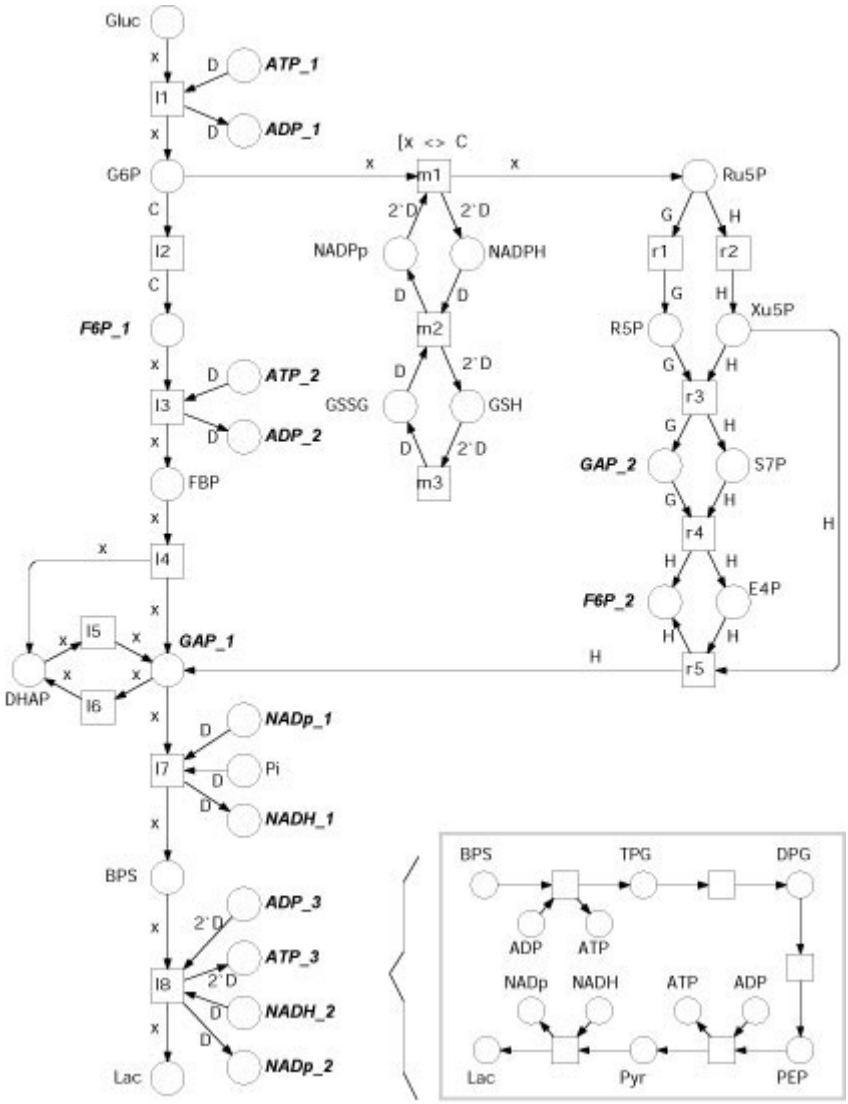

Fig.2. From the original P_{orig} to the high-level net P.

concomitant production of a relatively small amount of ATP. Then, pyruvate can be converted into lactate. The version chosen for this paper is that one for erythrocytes [see Stryer, 1996]. In the Petri net $\mathcal{P}$ (Figure 2), the *GP* consists of the reactions $l1$ to $l8$.

The *pentose phosphate pathway* (*PPP*), also called *hexose monophosphate pathway*, again starts with glucose and produces NADPH and ribose-5-phosphate (R5P) which then is transformed into glyceraldehyde-3-phosphate (GAP) and fructose-6-phosphate (F6P) and thus flows into the *GP*. In Figure 2, the *PPP* consists of the reactions $l1$, $m1$ to $m3$, $r1$ to $r5$, and $l3$ to $l8$. From now on we use acronyms for most metabolic substances; their full names are listed in the Abbreviations appendix.

It shall be noted that molecules like ADP, ATP, NAD^+, P_i etc. play a somewhat special role in metabolic networks. They are called *ubiquitous* because they are found in sufficiently large amounts in the cell. For ease of distinction, the remaining substances from Gluc to Lac shall be named *primary*. When talking about reactions in the following, the involved ubiquitous molecules commonly are not mentioned

because the primary substances are those that characterize the reactions and hence are of particular interest. Whereas the *GP* generates primarily ATP with glucose as a fuel, the *PPP* generates NADPH, which serves as electron donor for biosyntheses in cells. The interplay of the glycolytic and pentose phosphate pathways enables the levels of NADPH, ATP, and building blocks for biosyntheses, such as R5P and Pyr, to be continously adjusted to meet cellular needs. This interplay is quite complex, even in its somewhat simplified version that shall be discussed in this paper. In Reddy *et al.*, 1996, this pathway has already been modeled as a low-level Petri net (place/transition net) and then - qualitatively - analyzed by means of the well known linear algebraic methods. The analysis in Reddy *et al.*, 1996, is not completely correct, and it yields neither a *full* S-invariant (i. e., comprising all primary substances from the sources to the sinks) nor a non-trivial T-invariant. Hence it reveals some deficiencies that we will overcome by switching to high-level (colored) net models. Additionally, our models allow not only to be analyzed but also to be executed (simulated).

The construction of the low-level Petri net starts with a modest simplification of the original pathway model. The left branch of Figure 2 represents the *GP*. Its second part is a strictly linear path, transforming BPS into Lac via TPG, BPG, PEP, and Pyr (depicted in the small box at the lower right hand of Figure 2). This path can and shall be reduced to just one super-transition *l*8 (with the appropriate connections to ADP, ATP, NADH, and NAD^+) without altering the crucial properties of the net. In Design/CPN, such a node is called a *substitution transition*: *l*8 stands for and equivalently represents the mentioned linear path. With this modification and disregarding for a moment the guards and replacing all arc labels by '1', we get the place/transition net $\mathcal{P}$ of Figure 2, which is identical to the net found in Reddy *et al.*, 1996.

In organisms, the amount of the molecules is high enough to tolerate transient deviations from the theoretically postulated equilibrium concentrations. In the long run, a steady state is approximated: according to the kinetic equations, those reactions with higher reactant (lower product) concentrations are preferred to those with lower reactant concentrations (or higher product concentrations, resp.). In contrast, a qualitative analysis is confronted with small, even minimal amounts of molecules for any substrate.

The crucial point of using colored instead of low-level Petri nets is the following[1]. Applying higher-level places allows to discriminate between different molecules of the same metabolite via their identifiers (*colors*) C, D, F, ...[2]. This enables the designer to separate different branches of the compound pathway and to distinguish among molecules on the same place according to their origin and destination reaction. Although this distinction is not motivated by the biochemical reality (where molecules of the same substrate are identical), it increases the potential of the qualitative analysis we have in mind (see section "Steady state pathways, elementary modes"). As will be shown in section "Defects, effects and invariants", by choosing appropriate token colors we often get valuable information about the completeness and

[1] Legend for the Design/CPN nets in this paper:
- All places have the colorset , CS = C, D, F, G, H, G', H'.
- The underlined inscription 1 `D inside the place StartEnd denotes its initial marking.
- A place name in italics denotes a fusion place. All members of a fusion set are treated as the same place. Their names are numbered consecutively.
- The places for NAD+, NADP+, Pi, CO2, H2O are named NADp, NADPp, Pi, CO2, H2O respectively.
- A term in brackets [ ] is a guard (boolean expression) of the transition. If the value of the guard is true the transition may be enabled, if false it cannot.
- A (dashed) transition t' denotes the reverse counterpart of the reversible reaction t (see Figure 1).

[2] Note that the (theoretical) distinction by colors applies to chemically identical molecules, e. g., a token C on place G6P is distinguished from a token G on G6P. On the other hand, the substance that a molecule represents is unambiguously determined by the place it belongs to. Hence, a token C on G6P represents a different substance than a C on F6P.

feasability of the chosen model. Moreover, it is a prerequisite for executing such models properly, e. g., without running into unexpected deadlocks or the like. We have simulated all models in this paper, clearly not to get new results about their kinetics but mainly to gain confidence in the chosen color specifications.

What is the strategy of attributing colors to the tokens (molecules) along a given pathway model? Starting with a primary source substance of the pathway, we look for *conflicts* on the way to the sink(s). By definition, p is a conflict place if it has more than one output transition, and all are enabled if p carries a suitable token. If one of these alternative transitions occurs, all remaining transitions are disabled. In our context, a conflict would cause no harm as long as all but one alternative paths starting at p would end up again at this p without any lasting marking change. In general however, this is not the case. When looking at the metabolism in one specific organism, alternative metabolic paths most often result in different metabolic overall reactions. Therefore, they shall be discriminated and must not be combined deliberately. This discrimination is performed by attributing different identifiers to the molecules and by additionally blocking certain transitions for particular molecules (using guards).

This shall be demonstrated by use of the sample net $\mathcal{P}$, treated again as a low-level net by disregarding all arc labels and guards. Starting with the source Gluc, the first conflict is encountered at G6P which can be the reactant of either the reaction $l2$ or $m1$. A G6P-molecule with destination $l2$ gets a color, say C, and that one with destination $m1$ gets a different color (to be decided upon later). The guard $[x \neq C]$ prevents a C-token to be consumed by $m1$. Proceeding downwards the *GP* we examine *F6P_1*}, a fusion place. As *F6P_2*}, on the right-hand side, has no outgoing arc, a conflict does not exist. The next conflict on the way down is found at GAP (the fused *GAP_1* and *GAP_2*), a conflict among the three transitions $l6$, $l7$, and $r4$. The reaction $l6$ is trivial, as the loop *GAP_1* → $l6$ → DHAP → $l5$ → *GAP_1* returns the token to *GAP_1* without affecting any other places. The conflict between $l7$ and $r4$ is difficult to discuss at this moment without knowledge about the situation in the *PPP* at *GAP_2*}. We postpone it to the end of this paragraph. Instead, the path from G6P into the *PPP*, in the middle and right part of the figure, shall be inspected. Choosing a separate color F for the molecules of the middle part is not mandatory because there is no conflict here; it is just a matter of taste. The next conflict occurs at Ru5P with the choice to continue via $r1$ or $r2$. The colors of these two molecules must be different from each other and from C; we choose G and H. The last conflict at GAP is the postponed one. However, because the color G has been maintained from Ru5P via R5P until *GAP_2* and the tokens on *GAP_1* have the (different) color D, their distinction is accomplished already: the G-molecules are removed by reaction $r4$ and the D-molecules by $l7$. The resulting model is again $\mathcal{P}$, but now regarded as a colored net by including the arc labels and transition guards of Figure 2.

DEFECTS, EFFECTS AND INVARIANTS

The following calculations were made by use of an experimental software package SY, written by H. Genrich in Standard ML, for the symbolic analysis of colored Petri nets[3]. This package supports symbolic calculations based on the incidence matrix of an executable colored net in Design/CPN. It inspects and adopts the internal tables produced by the Design/CPN simulator for the graphical model and its data base. It allows, among others, to form symbolic dot products and matrix products, and to apply useful reduction rules and different formats for presenting the results.

[3] This package will be made available soon. If interested, please contact monika.heiner@informatik.tu-cottbus.de.

Defects and S-invariants

We start with looking for S-invariants in the net $\mathcal{P}$ of Figure 2. Obviously, there are four pairs of ubiquitous substrates which, if produced or consumed by a reaction, are transformed into each other, namely (ADP, ATP), ($NADP^+$, NADPH), (2 GSSG, GSH), and (NAD^+, NADH). This is verified by applying the function DEFECT of the package SY to the four S-vectors

[(ADP, _ ↦ 1`D), (ATP, _ ↦ 1`D)], [($NADP^+$, _ ↦ 1`D), (NADPH, _ ↦ 1`D)],

[(GSSG, _ ↦ 2`D), (GSH, _ ↦ 1`D)], and [(NAD^+ , _ ↦ 1`D), (NADH, _ ↦ 1`D)].

Doing this yields the null defect in all cases, which means that the S-vectors above constitute S-invariants[4].

Clearly, it is of much greater importance to deal with the *full* set of all primary (non-ubiquitous) substances. In steady state, there should exist an S-invariant comprising that *full primary set*. To find out the weight factors of a *full* S-vector, i. e., covering this set, we proceed step by step. First we observe that each molecule of FBP is transformed into two GAP-molecules by the reactions *l*4 and *l*5. We conclude that in any S-vector, to finally become an invariant, the place markings (number of molecules) of the glycolysis pathway *GP* from Gluc unto FBP must get a weight factor twice as high as GAP and the following places down to Lac. Trying to adopt the same principle to the pentose phosphate pathway *PPP*, however, leads to a non-null defect.

To be more specific, the simulation of $\mathcal{P}$ and also its T-invariants computed in the next subsection show that, starting with 3 Gluc molecules, the *GP* produces 6, and the *PPP* 5 Lac molecules. Because these alternative paths share the metabolites G6P, F6P, FBP, GAP, and BPS, it is not possible to find integer weight factors for these substances to make the full S-vector an invariant. We conclude that it is impossible to find a full S-invariant, in model $\mathcal{P}_{rev}$, with 'standard' means like low-level Petri nets or METATOOL [Pfeiffer *et al.*, 1999]. Hence, neither in Reddy *et al.*, 1996, nor in Schuster *et al.*, 1996, a full S-invariant has been reported.

Using high-level nets with individual tokens offers the possibility to distinguish the mentioned metabolites according to the paths along which they are produced and consumed. Constructing the desired S-invariant (not shown), we have to choose the weight 2 for all metabolites from Gluc unto FBP and E4P, irrespectively of the path on which they occur. For GAP, a threefold distinction has to be made: GAP-molecules produced by *r*3 or *r*4' (*x* = G) or by *r*5 (*x* = H) get the weight 2, whereas those produced by *l*4 or *l*5 (*x* = D) get the weight 1. And this distinction is kept also for BPS and Lac.

The result then is an S-invariant which, however, lacks a sensible biochemical interpretation because it is impossible to distinguish molecules of the same substance in organisms. Anyhow, this invariant lets us conclude that an essential product must be missing in the model of the *PPP*. Inspecting the *PPP* more carefully, shows that this product is the CO_2-molecule produced by the reaction complex *m*1:

$G6P + 2\ NADP^+ + H_2O \rightarrow Ru5P + 2\ NADPH + 2\ H^+ + CO_2$.

This means that the model $\mathcal{P}$ (and that of Reddy *et al.*, 1996) have to be revised for our purposes. Introducing both CO_2 and H_2O into the model means to add an input place H_2O and an output place CO_2 to *m*1,

[4] Adopting the definitions at the beginning of section "Steady state pathways, elementary modes" and the conventions of the package SY in a simplified version, an S-vector is written as a list of pairs (place *si*, distribution of *si*), where the second element - in our case - degenerates to a mapping of a color variable (or the don't-care symbol " _ " in case of a constant) to a linear combination over the standard colorset *CS* (cf. footnote 1).
The defect of an S-vector σ, computed symbolically by the function DEFECT, is given as a list of members *t*: *lico*(*CS*), in which *lico*(*CS*) denotes a linear combination of tokens that has to be added to an input or output place of transition *t* to make σ an S-invariant.
Dealing with the syntactic details of SY is far beyond the scope of this paper.

and an output place H_2O to $l8$ (the latter one originating in the H_2O produced by the reaction DPG → PEP + H_2O). Of course, also places for H^+ could be added to the model, but we decided to refrain from this in order to keep the readability of the model.

For this augmented model, a stepwise construction leads to the following full S-invariant (in a drastically simplified notation, just writing i instead of _ → i`D):

σ_C = [(Gluc, 6), (G6P, 6), (F6P, 6), (FBP, 6), (CO2, 1), (Ru5P, 5), (R5P, 5), (Xu5P, 5), (S7P, 7), (E4P, 4), (DHAP, 3), (GAP, 3), (BPS, 3), (Lac, 3)].

An inspection of σ_C reveals that the integer weight factor of any substance equals the number of C-atoms bound in it. Thus the S-invariant σ_C expresses the conservation rule that the sum of C-atoms bound by all involved substrates is constant. And this clearly represents a sensible biochemical interpretation.

Next we compute an S-invariant concerning the number of all O-atoms. In this case we obviously have to include also H_2O which, of course, did not appear in σ_C. We get

σ_O = [(Gluc, 6), (G6P, 6), (F6P, 6), (FBP, 6), (CO2, 2), (H2O, 1), (Ru5P, 5), (R5P, 5), (Xu5P, 5), (S7P, 7), (E4P, 4), (DHAP, 3), (GAP, 3), (P_i, 1), (BPS, 4), (Lac, 3)].

σ_O represents the conservation rule that the number of all O-atoms in the pathway is constant.

If looking for the P-atoms (or phosphate groups) we cannot expect to get a *full* S-invariant, as some of the primary substances do not contain a P. The S-vector

σ_P = [(ADP, 2), (ATP, 3), (G6P, 1), (F6P, 1), (FBP, 2), (Ru5P, 1), (R5P, 1), (Xu5P, 1), (S7P, 1), (E4P, 1), (DHAP, 1), (GAP, 1), (P_i, 1), (BPS, 2)]

is a *partial* S-invariant saying that the total number of P-atoms is constant. Moreover, σ_P is already reported in Reddy *et al.*, 1996. This is due to the fact that it contains neither CO_2 nor H_2O which are missing in their net model, as we know.

Finally, it should be mentioned that we also computed a full S-invariant concerning the sum of H-atoms, using a model that additionally includes all H^+ -ions (not shown).

Effects and T-invariants

T-vectors and T-invariants of a Petri net describe processes. This means that we have to take into account that reversible reactions may run in the backward direction.

As mentioned in section "Steady state pathways, elementary modes", in case of a reversible transition t, we add its complementary transition t' to the net. We start with the sample net model $\mathcal{P}$ (Figure 2), augmented by the places for CO_2 and H_2O. The reactions $l1$, $l3$, and $m1$ can be treated as irreversible, because we want to consider the *GP* and *PPP*, but not the gluconeogenesis. The linear path from BPS to Lac, replaced by the substitution transition $l8$, contains the irreversible reaction from PEP to Pyr; thus $l8$ is also treated as irreversible. Hence we introduce the new complementary transitions $l2'$, $l4'$, $l5'$, $l7'$, and $r1'$ to $r5'$ to the augmented net $\mathcal{P}$ and thus obtain the model $\mathcal{P}_{rev}$ in Figure 3 [5].

[5] Note 1. Transition $l6$ in the model $\mathcal{P}$ is identical to $l5'$ in $\mathcal{P}_{rev}$.
Note 2. $m2$ and $m3$ are treated as irreversible as they merely restore the consumed $NADP^+$ -molecules.
Note 3. The S-invariants of $\mathcal{P}_{rev}$ are identical to those computed in subsection "Defects and S-invariants" for $\mathcal{P}$.

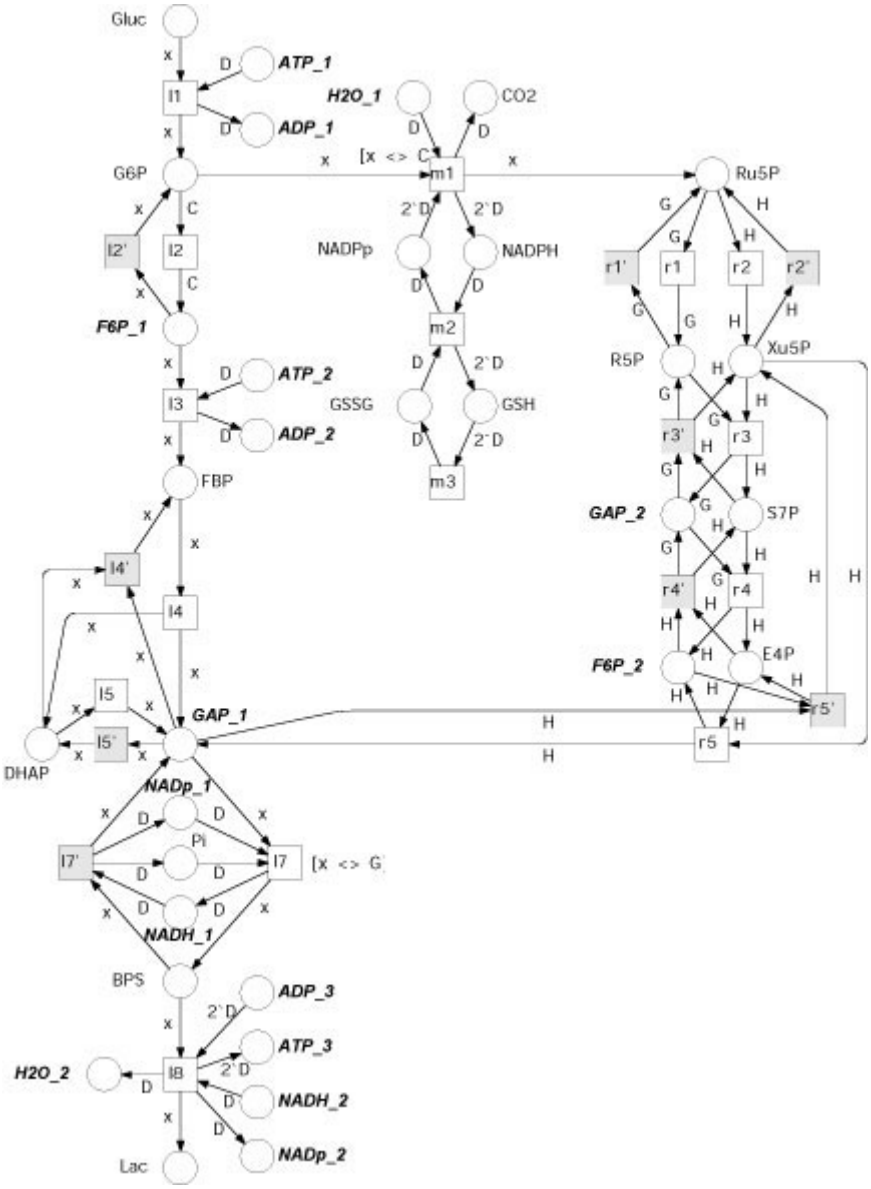

Fig.3. The Petri net $\mathcal{P}_{rev}$ with reversible reactions.

The introduction of the reverse transitions in $\mathcal{P}_{rev}$ may entail additional critical conflicts which have to be resolved when dealing with T-vectors and simulation. We observe:

- *l*4', *l*5', *l*7', and *r*1' to *r*5' merely create uncritical loops and can be deleted,
- *l*2' must not appear in the steady state *GP*.

Hence, *l*2' must be prevented from occurring for tokens x = C. In the course of *PPP*, one G- and two H-molecules on G6P are transformed into one H-token on *GAP_1* (via *r*5) and two H-tokens on *F6P_2* (via *r*4 and *r*5).

For the latter H-tokens there are three possibilities to be processed further:

1. both move to FBP via *l*3 and continue on the 'normal' way to Lac,
2. both move to G6P via *l*2' and thus regenerate the initial two H-tokens there,

3. one of them is moved by *l*3 to FBP, and the second one by *l*2' to G6P. This case is a combination of glycolysis and *gluconeogenesis*, which cannot occur in steady state.

To distinguish the 'normal' *PPP* path (1) from the new reaction path (2) we have to introduce new token-colors, G′ and H′ say, for (2). With exception of *l*2', the processing of G′ and H′ is identical to that of G and H; therefore in the *PPP*-branch, the token color instances (identifiers) are replaced by the variables *x* (for G or G′) and *y* (for H or H′), and - a technicality - appropriate guards are attributed to the reactions *r*1 to *r*4. Finally, because the molecules moved onto F6P by the glycolysis resp. the 'normal' *PPP* are C resp. H, the reaction *l*2' may be enabled only for H′-molecules. Hence the arc pointing to *l*2' gets the label H′ and the reaction *l*3 gets the guard [*x* <> *H′*].

This leads to $\mathcal{P}^*$ in Figure 4 which is appropriate both for computing T-invariants and for simulation because all unreasonable processes and cyclic loops have been excluded.

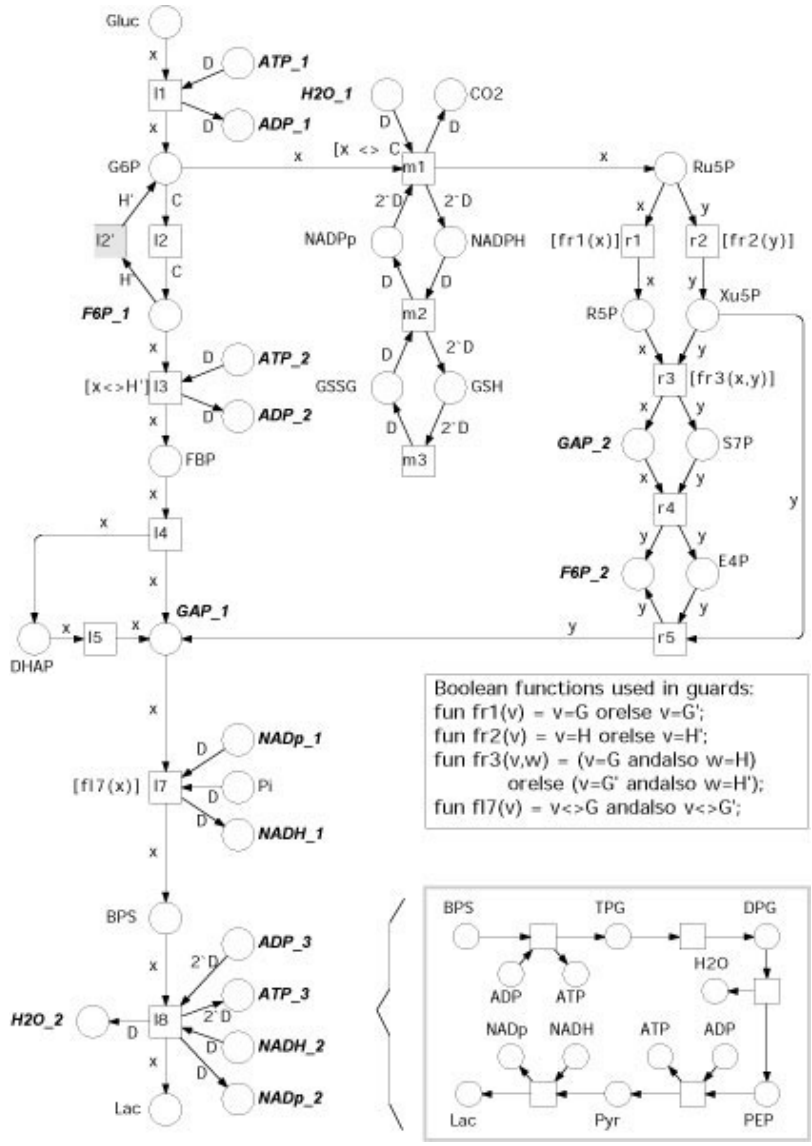

Fig.4. The net $\mathcal{P}^*$ with three flux modes.

Looking for T-invariants which describe the feasible processes in the net we stay, for a moment, with $\mathcal{P}_{rev}$. Clearly, for each reversible reaction t and the reverse reaction t', the vector $[(t, 1), (t', 1)]$ is a T-invariant. But these T-invariants lack a sensible biochemical interpretation: a reversible reaction occurring permanently in both directions does not make sense. For this reason all reverse reactions except *l*2' have been omitted at the end of section "Models of the glycolysis and pentose phosphate pathway". On the other hand, we observe that *l*2' gives rise to a process that is different from both the *GP* and the *PPP*, namely the gluconeogenesis. The tokens of that extra process got the identifiers G′ and H′, leading to the net model $\mathcal{P}^*$.

As with the S-vectors, also the T-vectors shall be established stepwise, i. e., not automatically but systematically. Apart from being able to see, from the *effects* computed at each step, the weight factor(s) of the transition(s) to be added successively to the T-vector, there is still another advantage of this approach. If we proceed along the causal chain of transitions, i. e., at each step selecting a subsequent transition which is enabled, we can compile knowledge about the amount of molecules which are needed *in course of* the run from Gluc to Lac and which have to be restored later during the run or after its end. This information is not provided by the effect of the complete T-vector: it only shows the *overall* effect of the vector.

When looking for the possible processes in the *GP*/*PPP* system $\mathcal{P}^*$ we soon find out that there are (at least) three sorts of processes (modes) that can be run independently from each other. Therefore we can attribute to each of them a characteristic parameter by which the weighted vector elements are multiplied additionally, namely

– *gly* for the glycolysis pathway,
– *hex* for the pentose (or hexose mono-) phosphate pathway, and
– *rev* for the pathway including the reverse reaction *l*2'.

During the stepwise construction of the T-vector(s) we gather, on the one hand, information about those molecules that are needed at the beginning or in the course of a run to reach its end at Lac. These molecules and their amounts are:

Consumed substances			
ADP:	$(2\cdot\ gly + 5 \cdot hex + rev)$`D	ATP:	$(2 \cdot gly + 5 \cdot hex + rev)$`D
F6P:	$2 \cdot rev$ `H′	Gluc:	gly `C + $hex$ `G + $2 \cdot hex$ `H + $rev$`G′
GSSG:	$(6 \cdot hex + 6 \cdot rev)$`D	H_2O:	$(3 \cdot hex + 3 \cdot rev)$`D
NAD^+:	$(2 \cdot gly + 5 \cdot hex + rev)$`D	$NADP^+$:	$(6 \cdot hex + 6 \cdot rev)$`D
P_i:	$(2 \cdot gly + 5 \cdot hex + rev)$`D		

On the other hand, we gather information about those molecules that are provided during or at the end of the full run. These molecules and their amounts are:

Produced substances			
ATP:	$(4 \cdot gly + 10 \cdot hex + 2 \cdot rev)$`D	CO_2:	$(3 \cdot hex + 3 \cdot rev)$`D
F6P:	$2 \cdot rev$`H′	GSSG:	$(6 \cdot hex + 6 \cdot rev)$`D
H_2O:	$(2 \cdot gly + 5 \cdot hex + rev)$`D	Lac:	$2 \cdot gly$`C + $5 \cdot hex$`H + rev `H′
NAD^+:	$(2 \cdot gly + 5 \cdot hex + rev)$`D	$NADP^+$:	$(6 \cdot hex + 6 \cdot rev)$`D

The final result of the construction is the complete parameterized T-vector [6].

$$
\begin{aligned}
\tau' = [\ & (l1, [\ (1, gly`(x \mapsto C)), (1, hex`(x \mapsto G)),\\
& \quad (2, hex`(x \mapsto H)), (1, rev`(x \mapsto G'))\ ]),\\
& (l2, [\ (1, gly`(\_))\ ]),\\
& (l3, [\ (1, gly`(x \mapsto C)), (2, hex`(x \mapsto H))\ ]),\\
& (l4, [\ (1, (gly + 2 \cdot hex)`(\_))\ ]),\\
& (l5, [\ (1, (gly+2 \cdot hex)`(\_))\ ]),\\
& (l7, [\ (1, (2 \cdot gly + 5 \cdot hex + rev)`(\_))\ ]),\\
& (l8, [\ (1, (2 \cdot gly +5 \cdot hex + rev)`(\_))\ ]),\\
& (m1,[\ (1, hex`(x \mapsto G)), (1, rev`(x \mapsto G')),\\
& \quad (2, hex`(x \mapsto H)), (2, rev`(x \mapsto H'))\ ]),\\
& (m2,[\ (6, (hex + rev)`(\_))\ ]),\\
& (m3,[\ (6, (hex + rev)`(\_))\ ]),\\
& (r1, [\ (1, hex`(x \mapsto G)), (1, rev`(x \mapsto G'))\ ]),\\
& (r2, [\ (2, hex`(y \mapsto H)), (2, rev`(y \mapsto H'))\ ]),\\
& (r3, [\ (1, hex`((x,y) \mapsto (G,H))), (1, rev`((x,y) \mapsto (G',H')))\ ]),\\
& (r4, [\ (1, hex`((x,y) \mapsto (G,H))), (1, rev`((x,y) \mapsto (G',H')))\ ]),\\
& (r5, [\ (1, hex`(y \mapsto H)), (1, rev`(y \mapsto H'))\ ]),\\
& (l2', [\ (2, rev`(\_))\ ])\ ].
\end{aligned}
$$

The T-vector τ' represents three vectors, one for each parameter, which correspond to the expected three elementary modes. They are identical to those computed by S. Schuster, using METATOOL. Applying the function EFFECT from the package SY to τ' yields its effect, which equals the difference *Produced substances* minus *Consumed substances*,

$$
\begin{aligned}
&\text{ADP:} && -2 \cdot gly`D - 5 \cdot hex`D - rev`D,\\
&\text{ATP:} && 2 \cdot gly`D + 5 \cdot hex`D + rev`D,\\
&CO_2\text{:} && 3 \cdot hex`D + 3 \cdot rev`D,\\
&\text{Gluc:} && -gly`C - hex`G - 2 \cdot hex`H - rev`G',\\
&H_2O\text{:} && 2 \cdot gly`D + 2 \cdot hex`D - 2 \cdot rev`D,\\
&\text{Lac:} && 2 \cdot gly`C + 5 \cdot hex`H + rev`H',\\
&P_i\text{:} && -2 \cdot gly`D - 5 \cdot hex`D - rev`D.
\end{aligned}
$$

[6] Adopting the conventions of the package SY in a simplified version, T-vectors are written as lists of pairs (transition name, list of weighted substitutions). A weighted substitution consists of an integer weight, followed by an integer parameter, a multiplication sign " ` " and a substitution in parentheses ( ). A substitution, represented by " $\mapsto$ ", indicates which variable(s) of the arc labels have to be substituted by which color(s). If the arc labels are constant colors the don't-care symbol (_) is used.
The effect of a T-vector τ is presented as a list of constructs *s*: *lico*(*CS*), where *lico*(*CS*) denotes a linear combination of tokens that has to be added to (or subtracted from) place *s* to make τ a T-invariant.

Neglecting the token colors, this leads to the parameterized equation for the effect of τ'
$(2 \cdot gly + 5 \cdot hex + rev)$`ADP + $(gly + 3 \cdot hex + rev)$`Gluc + $(2 \cdot gly + 5 \cdot hex + rev)$`$P_i$ + $2 \cdot hex$`$H_2O$ = $(2 \cdot gly + 5 \cdot hex + rev)$`ATP + $(2 \cdot gly + 5 \cdot hex + rev)$`Lac + $(3 \cdot hex + 3 \cdot rev)$`$CO_2$ + $(2 \cdot gly + 2 \cdot hex)$`$H_2O$
yielding the three *overall reaction* equations for the elementary modes

	Parameters	Overall reaction
(G)	$gly = 1$	2 ADP + Gluc + 2 P_i = 2 ATP + 2 Lac + 2 H_2O
(P)	$hex = 1$	5 ADP + 3 Gluc + 5 P_i = 5 ATP + 5 Lac + 3 CO_2 + 2 H_2O
(R)	$rev = 1$	ADP + Gluc + P_i + 2 H_2O = ATP + Lac + 3 CO_2

T-invariants describe processes in a Petri net which restore the marking with which they started and thus can be executed cyclically. Clearly, τ' is *not* a T-invariant. Because in $\mathcal{P}^*$ all paths containing Gluc and Lac are not cyclic, *no* full T-vector *at all* can be a T-invariant. Therefore, we modify the model $\mathcal{P}^*$ by glueing it with a subnet $\mathcal{P}^{se}$ (depicted in Figure 5) that closes the cycle from Lac to Gluc. This subnet contains a place StartEnd, initially marked by a dummy token D, and the transitions s1 for starting and s2 for ending a cyclic run. The transitions $s1$ and $s2$ are intended to compensate the non-null effects. To this end, we connect appropriate substances of $\mathcal{P}^*$ (as fusion places) with $s1$ and/or $s2$. At this stage, one of the advantages of the stepwise construction of the T-vector τ' becomes clear. The tables *Consumed Substances* resp. *Produced Substances*, derived above, exactly inform about those substances and their amounts that have to be provided by $s1$ resp. removed by $s2$ to arrive at a (parameterized) T-invariant.

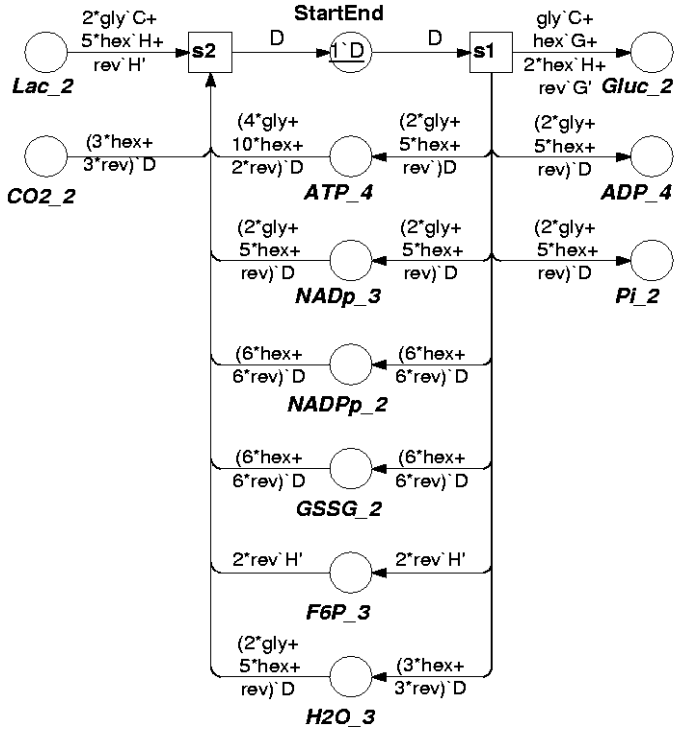

Fig.5. The subnet $\mathcal{P}^{se}$ completing $\mathcal{P}^*$ to form a cycle.

Combining this subnet $\mathcal{P}^{se}$ with $\mathcal{P}^*$ by means of the fusion places, yields the cyclic net model that we aimed at. Let τ denote the T-vector achieved by adding the elements
($s1$, [(1, (_))]) and ($s2$, [(1, (_))]) to τ'. Then this T-vector τ has no effect and hence is a parameterized T-invariant.

The minimal T-invariants derived from τ by setting one of the parameters *gly*, *hex*, or *rev* to 1 (and the remaining two to 0) are, in a short-hand notation,

$\tau_G =$ [($l1$, C), ($l2$, D), ($l3$, C), ($l4$, D), ($l5$, D), ($l7$, 2 · D), ($l8$, 2 · D), ($s1$, D), ($s2$, D)],

$\tau_P =$ [($l1$, G + 2 · H), ($l3$, 2 · H), ($l4$, 2 · H), ($l5$, 2 · H), ($l7$, 5 · H), ($l8$, 5 · H),
($m1$, G + 2 · H), ($m2$, 6 · D), ($m3$, 6 · D),
($r1$, G), ($r2$, 2 · H), ($r3$, (G,H)), ($r4$, (G,H)), ($r5$, H), ($s1$, D), ($s2$, D)],

$\tau_R =$ [($l1$, G′), ($l2'$, 2 · H′), ($l7$, H′), ($l8$, H′), ($m1$, G′ + 2 · H′), ($m2$, 6 · D), ($m3$, 6 · D),
($r1$, G′), ($r2$, 2 · H′), ($r3$, (G′,H′)), ($r4$, (G′,H′)), ($r5$, H′), ($s1$, D), ($s2$, D)].

The three T-invariants τ_G, τ_P, τ_R are linearly independent, hence form a basis.

Biochemical evaluation of the T-invariants

The software package METATOOL [Pfeiffer *et al.*, 1999] allows to compute the *elementary (flux) modes* (corresponding to the 'non-cyclic portions' of the minimal T-invariants) of a pathway. For each mode, it computes (1) the T-vector, determining which reactions have to occur how often to proceed from the sources to the sinks, and (2) the overall reaction equation.

With the colored Petri net approach and applying the package SY, we get additional information not only about the T-invariants but also about the dynamics of the system. The symbolic treatment of the T-vectors yields as one crucial result the marking (amount of molecules), needed at the beginning and provided by the starting transition $s1$, to run the system without deadlock from its source to the sink. This initial marking is 'appropriate' because it is the minimum amount of molecules necessary for a simulation. Moreover, the stepwise construction of the symbolic parameterized T-invariants yields knowledge not only about the frequency of transition occurrences (during a run along the invariant) but also about the partial order in which these transitions have to occur.

An interesting question arises concerning the independence of the three T-invariants. Theoretically, they are linearly independent because the transitions $l2$ and $l2'$ are treated as not being related to each other. If however $l2$ and - $l2'$ are identified, the T-vectors get linearly dependent. This corresponds to the observation that the overall reactions (G), (P), (R) are related to each other by the equation (P) = 2 · (G) + (R).

The problem, however, lies in the fact that a steady state process including both a reaction ($l2$) and its *rev*erse ($l2'$) is biochemically not feasible. And on the other hand, T-vectors with negative elements cannot be T-invariants according to the definition given in section "Steady state pathways, elementary modes".

The construction of the compound net $\mathcal{P}^*$ can also be looked at from a different perspective, throwing more light on the nature of the token colors and the conflicts. Let us discuss the three independent modes identified in the previous subsection "Effects and T-invariants", as separate net models. They are depicted in a simplified version as Figure 6, omitting all ubiquitous molecules and the 'uncritical' reactions $m2$ and $m3$.

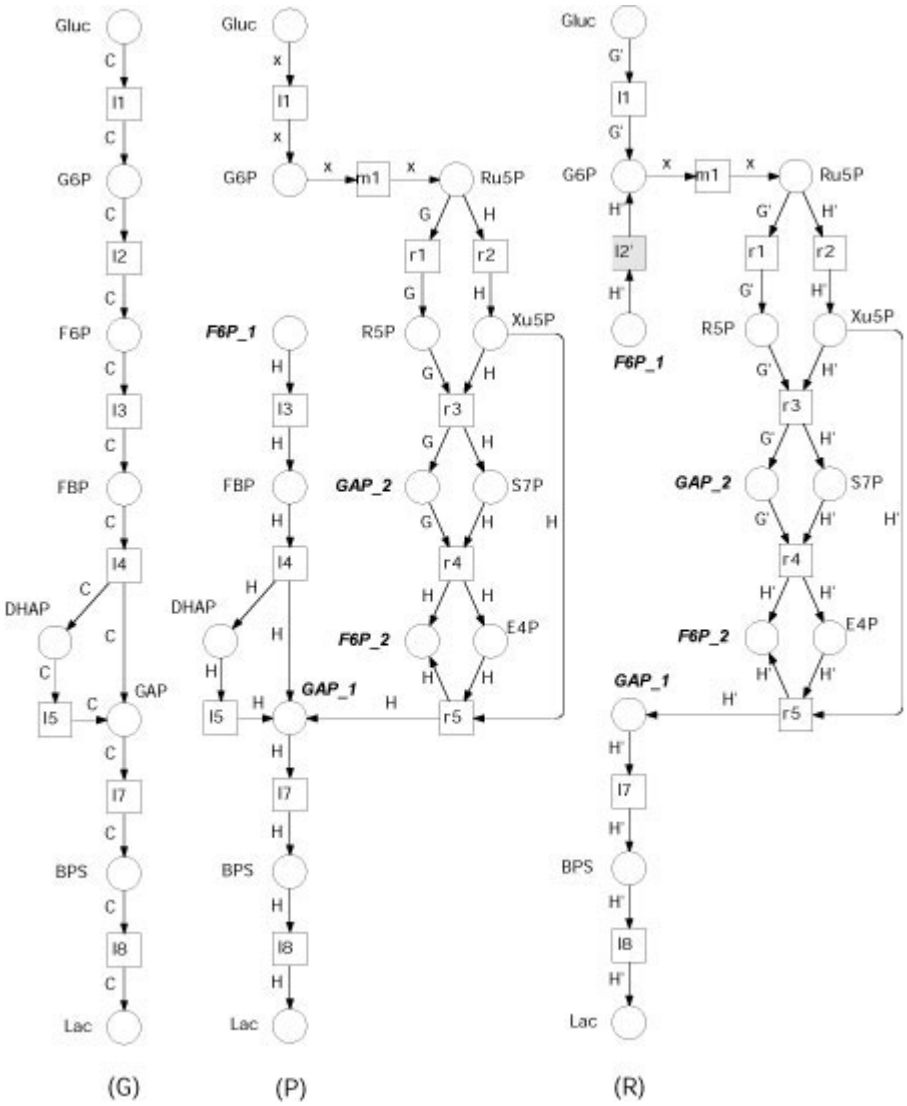

Fig.6. The (simplified) three modes of $\mathcal{P}^*$.

The first mode (G), glycolysis, contains no conflict. So, only one token color, C, is needed. The second mode (P) has two internal conflicts at Ru5P and GAP which are decided by use of the two colors G and H. The third mode (R) contains the same two conflicts as (P), now solved by G′ and H′.

These 'mode specific' conflicts describe (model) situations as happening in reality, with a great number of molecules of every substance involved. From the definition of steady state it follows that, sloppy speaking, no molecule inserted by the source may get stuck on its way to the sink. If it would, the concentration of an intermediate substance would be increased, contradicting the definition. Looking at the right hand branch of (P) in Figure 6, the tokens entering that branch at Ru5P can leave it only as F6P- or GAP-molecules by means of *r*4 and *r*5. The reaction *r*4 needs one G and one H, and *r*5 one additional H. The one G or two H tokens, resp., can only be provided by *r*1 occurring once or by *r*2 occurring twice, respectively. In organisms, the molecules of one substance cannot be distinguished and cannot be forced to choose one out of more alternative paths. Yet, the transitory increase of a substance concentration leads to

a slowing down of reactions producing it and an acceleration of reactions consuming it. The opposite happens in case of a concentration decrease. So, in the long run, a relative occurrence ratio of 1:2 will be established among $r1$ and $r2$.

In contrast to the mode specific conflicts, the remaining ones are consequences of glueing the mode nets (G), (P), and (R) into one single model $\mathcal{P}^*$. As these three processes are independent from each other, performing linear independent T-invariants, the parameters *gly*, *hex*, *rev* can, in principle, be chosen arbitrarily. This implies that the relative frequency among this kind of conflicting reactions, for example $l2$ and $m1$, depends merely on the choice of the parameters and not on a biochemical law that would require a constant frequency ratio. In an organism, these reaction ratios are controlled mainly by the current needs of the cell, governing the relative activities of the respective enzymes.

The flow of G6P or Gluc depends on the need for NADPH, R5P, and ATP in the cell. Based on experimental observation, biochemists distinguish between four 'modes' (which we will call *T-modes*, in order to not mistake them for the elementary flux modes) of the combined *GP*/*PPP* [see Stryer, 1996]. We finish this section with a short discussion of these T-modes and their relationships to our results, neglecting again H^+.

T-Mode 1 is adopted when more R5P than NADPH is required, for example in rapidly dividing cells needing R5P for the synthesis of nucleotide precursors of DNA. Most of G6P is converted into F6P and GAP by the *GP* ($l2$, $l3$, $l4$). Transaldolase ($r4'$) and transketolase ($r3'$) then convert 2 F6P- and 1 GAP- into 3 R5P-molecules. The reaction reads

$$5\ \mathrm{G6P} + \mathrm{ATP} \rightarrow 6\ \mathrm{R5P} + \mathrm{ADP}.$$

First, we recognize that we have to return to the model $\mathcal{P}_{rev}$ which contains all *rev*erse reactions needed. Secondly, we observe that the process (reaction path) does not transform Gluc into Lac, hence, does not represent a full T-vector. As a consequence, the chosen token colors are no longer appropriate. Bearing this in mind, we construct the T-vector

$$[\ (l2,\ [\ (5,\ (\_))\ ]),\ (l3,\ [\ (1,\ (\_)\ ]),\ (l4,\ [\ (1,\ (\_)\ ]),\ (l5,\ [\ (1,\ (\_)\ ]),\ (r1,\ [\ (4,\ (\_)\ ]),$$
$$(r5',\ [\ (2,\ (\_)\ ]),\ (r4',\ [\ (2,\ (\_)\ ]),\ (r3',\ [\ (2,\ (\_)\ ]),\ (r2',\ [\ (4,\ (\_)\ ])\ ]$$

and compute its effect, yielding

$$\mathrm{ADP}: \mathrm{D},\ \mathrm{ATP}: -\mathrm{D},\quad \mathrm{F6P}: 5\ \grave{}\mathrm{C} - 2\ \grave{}\mathrm{H} - 3\ \grave{}\mathrm{x},\ \mathrm{G6P}: -5\ \grave{}\mathrm{C},$$
$$\mathrm{GAP}: 2\ \grave{}\mathrm{D} - 2\ \grave{}\mathrm{H},\ \mathrm{R5P}: 6\ \grave{}\mathrm{G},\quad \mathrm{Ru5P}: -4\ \grave{}\mathrm{G} + 4\ \grave{}\mathrm{H}.$$

By identifying all token colors with D, say, the effects for F6P, GAP, and Ru5P disappear, leading to the desired overall effect

$$\mathrm{ADP}: \mathrm{D},\quad \mathrm{ATP}: -\mathrm{D},\quad \mathrm{G6P}: -5\ \grave{}\mathrm{D},\quad \mathrm{R5P}: 6\ \grave{}\mathrm{D}$$

which exactly reflects the reaction formula above.

T-mode 2 is adopted when the needs for NADPH and R5P are balanced. Then the oxidative branch of the *PPP* is executed, converting G6P into NADPH and R5P via $m1$ and $r1$. The reaction formula is

$$\mathrm{G6P} + 2\ \mathrm{NADP^+} + \mathrm{H_2O} \rightarrow \mathrm{R5P} + 2\ \mathrm{NADPH} + \mathrm{CO_2}.$$

For the T-vector

$$[\ (m1,\ [\ (1,\ (x \mapsto \mathrm{G}))\ ]),\ (r1,\ [\ (1,\ (\_)\ ])]$$

we verify the corresponding effect

$$\mathrm{CO_2}: \mathrm{D},\ \mathrm{G6P}: -\mathrm{G},\ \mathrm{H_2O}: -\mathrm{D},\ \mathrm{NADPH}: 2\ \grave{}\mathrm{D},\ \mathrm{NADP^+}: -2\ \grave{}\mathrm{D},\ \mathrm{R5P}: \mathrm{G}.$$

T-modes 3 and 4 are adopted when much more NADPH than R5P is required.

T-mode 3 reads

$$G6P + 12\ NADP^+ + 7\ H_2O \rightarrow 6\ CO_2 + 12\ NADPH + P_i.$$

It includes, apart from $l4'$ and $l2'$, a reaction $l3^*$: FBP →F6P, catalyzed by fructose-1,6-biphosphatase, which is part of the gluconeogenesis and hence outside the scope of the *GP*/*PPP* system covered by the models of this paper.
Note that both $l3^*$ and $l3$: F6P →FBP are irreversible.

T-mode 4, according to Stryer, 1996, is characterized by the reaction formula

$$3\ G6P + 6\ NADP^+ + 5\ NAD^+ + 5\ P_i + 8\ ADP$$
$$\rightarrow 5\ Pyr + 3\ CO_2 + 6\ NADPH + 5\ NADH + 8\ ATP + 2\ H_2O.$$

If this process is expanded to start with Gluc (instead of G6P) and to end with Lac (instead of Pyr), the result corresponds precisely to the T-invariant τ_P derived in the previous subsection "Effects and T-invariants".

CONCLUSIONS

To our best knowledge, this paper is the first one to apply higher-level Petri nets to the design, qualitative analysis, and execution of metabolic steady state system models. Compared to low-level Petri nets and to algebraic methods and tools from biochemistry, this approach renders important new results about the invariants and the processes of (sufficiently complex) metabolic pathways. The crucial point of using high-level nets is the ability to discriminate metabolites, if necessary, according to their topological environment, i. e., the reaction chains in which they are involved. On this basis, models can be developped which can be simulated smoothly and can be subjected to a rigorous symbolic analysis. This has been demonstrated for the rather complex sample of the combined glycolysis and pentose phosphate pathways. Our main results are the following.

Firstly, some full S-invariant of the sample net were found that represent interesting, non-trivial preservation laws for the total amounts of certain atoms or molecules in the system. Additionally, their incremental construction may reveal inconsistencies or deficiencies of the examinated model.

Secondly, the elementary modes (and the corresponding T-invariants) and their overall reaction equations as computed by METATOOL have been verified. These three T-invariants have been represented as one parameterized vector. Moreover, not only the number of reaction occurrences related to a T-invariant, but also their partial order has been determined.

Thirdly, the sample net model can be simulated cyclically, restoring the initial system state at the end of each cycle, avoiding deadlocks, and respecting the inherent concurrency.

Fourthly, a biochemical interpretation of high-level Petri net models of steady state pathways and their invariants may enhance the understanding of metabolic processes.

A most interesting topic for further research is the question whether or to which extent the search for and the construction of S- and T-invariants can be automated. Moreover, the significance of (full) S-invariants and defects deserves an increased attention. Finally, the application of symbolic analysis to less understood metabolic systems is expected to lead to valuable new results.

ACKNOWLEDGEMENTS

We thank for the financial support by the BMBF (Federal Ministry of Education and Research of Germany), BCB project number 0312705D. Further we would like to thank H. Genrich and S. Schuster for fruitful discussions.

ABBREVIATIONS

Metabolites / Compounds			
ADP	Adenosine diphosphate	ATP	Adenosine triphosphate
BPS	1,3-Biphosphoglycerate	DHAP	Dihydroxyacetone phosphate
DPG	2-Phosphoglycerate	E4P	Erythose-4-phosphate
FBP	Fructose biphosphate	F6P	Fructose-6-phosphate
GAP	Glyceraldehyde-3-phosphate	Gluc	Glucose
GSH	Glutathione	GSSG	Glutathionedisulfide
G6P	Glucose-6-phosphate	Lac	Lactate
NADH	Nicotinamide adenine dinucleotide, reduced form		
NAD^+	NADp, Nicotinamide adenine dinucleotide, oxidized form		
NADPH	Nicotinamide adenine dinucleotide phosphate, reduced form		
$NADP^+$	NADPp, Nicotinamide adenine dinucleotide phosphate, oxidized form		
PEP	Phosphoenolpyruvate	P_i	Orthophosphate, ionic form
Pyr	Pyruvate	Ru5P	Ribulose-5-phosphate
R5P	Ribose-5-phosphate	S7P	Sedoheptulose-5-phosphate
TPG	3-Phosphoglycerate	Xu5P	Xylulose-5-phosphate

Correspondence between Petri net transitions and enzymatic reactions			
*l*1	Hexokinase	*l*2	Phosphoglucose isomerase
*l*3	Phosphofructokinase	*l*4	Aldolase
*l*5	Triosephosphate isomerase (forw.)	*l*6	Triosephosphate isomerase (backw.)
*l*7	GAP dehydrogenase		
*l*8	*Reaction path consisting of:* phosphoglycerate kinase, phosphoglycerate mutase, enolase, pyruvate kinase, and lactate dehydrogenase		
*m*1	G6P *oxidation reactions*	*m*2	Glutathione reductase
*m*3	Glutathione oxidation reaction	*r*1	Ribulose-5-phosphate isomerase
*r*2	Ribulose-5-phosphate epimerase	*r*3	Transketolase
*r*4	Transaldolase	*r*5	Transketolase

REFERENCES

- Design/CPN. http://www.daimi.au.dk/designCPN/
- Genrich, H. (2002). Dynamical Quantities in Net Systems. Formal Aspects of Computing **14**, 55-89.
- Genrich, H., Kueffner, R. and Voss, K. (2001). Executable Petri Net Models for the Analysis of Metabolic Pathways. Int. J. STTT **3**, 394-404.
- Goss, P. J. E. and Peccoud, J. (1998). Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. Proc. Natl. Acad. Sci. USA **95**, 6750-6755.
- Goss, P. J. E. and Peccoud, J. (1999). Analysis of the stabilizing effect of Rom on the genetic network controlling ColE1 plasmid replication. Pac. Symp. Biocomp. **4**, 65-76.
- Heinrich, R. and Schuster, S. (1996). The Regulation of Cellular Systems. Chapman & Hall, New York.
- Heinrich, R. and Schuster, S. (1998). The modeling of metabolic systems. Structure, control and optimality. BioSystems **47**, 61-77.
- Hofestaedt, R. (1994). A Petri Net Application of Metabolic Processes. Journal of System Analysis, modeling and Simulation **16**, 113-122.
- Hofestaedt, R. and Thelen, S.(1998). Quantitative Modeling of Biochemical Networks. In Silico Biol. **1**, 0006.
- Jensen, K. (1992-1997). Coloured Petri Nets. Basic Concepts, Analysis Methods and Practical Use. Vol. 1-3. Monographs in Theoretical Computer Science. Springer-Verlag, Berlin.
- Koch, I., Schuster, S. and Heiner, M. (2000). Using time-dependent Petri nets for the analysis of metabolic networks. *In:* DFG-Workshop: Informatikmethoden zur Analyse und Interpretation grosser genomischer Datenmengen, Hofestaedt, R., Lautenbach, K. and Lange, M. (eds.), Magdeburg, pp. 15-21.
- Matsuno, H., Doi, A., Nagasaki, M. and Miyano, S. (2000). Hybrid Petri net representation of gene regulatory network. Pac. Symp. Biocomput. **5**, 338-349.
- Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J. C., Montero, F. and Schuster, S. (1999). METATOOL: For studying metabolic networks. Bioinformatics **15**, 251-257.
- Reddy, V. N., Mavrovouniotis, M. L. and Liebman, M. N. (1993). Petri Net Representation in Metabolic Pathways. *In* Proc. First Intern. Conf. on Intelligent Systems for Molecular Biology, Hunter, L. *et al.* (please, complete list of editors!) (eds.), AAAI Press, Menlo Park, pp. 328-336
- Reddy, V. N., Liebman, M. N. and Mavrovouniotis, M. L. (1996). Qualitative analysis of biochemical reaction systems. Comput. Biol. Med. **26**, 9-24.
- Schuster, S. and Hilgetag, C. (1994). On elementary flux modes in biochemical reaction systems at steady state. J. Biol. Syst. **2**, 165-182.
- Schuster, S., Hilgetag, C., Woods, J. H. and Fell, D. A. (1996). Elementary Modes of Functioning in Biochemical Networks. *In*: Computation in Cellular and Molecular Biological Systems, Cuthbertson, R., Holcombe, M. and Paton, R. (eds.), World Scientific, Singapore, pp. 151-165.
- Schuster, S., Fell, D. A. and Dandekar, T. (2000a). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnol. **18**, 2000, 326-332.
- Schuster, S., Pfeiffer, T., Moldenhauer, F., Koch, I. and Dandekar, T. (2000b). Structural Analysis of metabolic Networks: Elementary Flux Modes, Analogy to Petri Nets, and Application to Mycoplasma pneumoniae. *In:* Bauer, E.-B., Rost, U., Stoye, J., Vingron, M. (eds.), Proc. Germ. Conf. Bioinf., Heidelberg, Logos Verlag Berlin, pp. 115-120.
- Stryer, L. (1996). Biochemistry. W. H. Freeman and Co., New York.