BRANDENBURG UNIVERSITY OF TECHNOLOGY AT COTTBUS

Faculty of Mathematics, Natural Sciences and Computer Science

Institute of Computer Science

COMPUTER SCIENCE REPORTS

Report 03/04

August 2004

QUALITATIVE PATH ANALYSIS OF METABOLIC PATHWAYS USING PETRI NETS FOR GENERIC MODELLING

THOMAS RUNGE

Computer Science Reports
Brandenburg University of Technology at Cottbus
ISSN: 1437-7969

Send requests to: BTU Cottbus
Institut für Informatik
Postfach 10 13 44
D-03013 Cottbus

Thomas Runge
thomas.runge@informatik.tu-cottbus.de,
http://www-dssz.informatik.tu-cottbus.de

Qualitative Path Analysis of Metabolic Pathways Using Petri Nets for Generic Modelling

Computer Science Reports
03/04
August 2004

Brandenburg University of Technology at Cottbus

Faculty of Mathematics, Natural Sciences and Computer Science

Institute of Computer Science

Computer Science Reports
Brandenburg University of Technology at Cottbus
Institute of Computer Science

Head of Institute:
Prof. Dr. Claus Lewerentz cl@informatik.tu-cottbus.de
BTU Cottbus
Institut für Informatik
Postfach 10 13 44
D-03013 Cottbus

Research Groups:	Headed by:
Computer Engineering	Prof. Dr. H. Th. Vierhaus
Computer Network and Communication Systems	Prof. Dr. H. König
Data Structures and Software Dependability	Prof. Dr. M. Heiner
Programming Languages and Compiler Construction	Prof. Dr. P. Bachmann
Software and Systems Engineering	Prof. Dr. C. Lewerentz
Theoretical Computer Science	Prof. Dr. B. von Braunmühl
Graphics Systems	Prof. Dr. W. Kurth
Systems	Prof. Dr. R. Kraemer
Distributed Systems and Operating Systems	Prof. Dr. J. Nolte

CR Subject Classification (1998): D.2.2, I.6.3, I.6.5, I.6.4, J.3

Printing and Binding: BTU Cottbus

ISSN: 1437-7969

Qualitative Path Analysis of Metabolic Pathways Using Petri Nets for Generic Modelling

Thomas Runge[1,2]
[1]Brandenburg University of Technology Cottbus, Department of Computer Science,
chair of data structures and software dependability
Postbox 10 13 44, 03013 Cottbus, Germany
thomas.runge@informatik.tu-cottbus.de
phone: +49 355 69 3885
fax: +49 355 69 3830
[2]Technical University of Applied Sciences Berlin, Department of Bioinformatics,
Seestrasse 64, 13347 Berlin, Germany

Abstract

Computer aided analysis possibilities are necessary to improve the understanding of the complex biochemical processes of a cell. The results of kinetic models are often non-reliable on account of the lack of reliable data. Therefore, other supplementary methods are indispensable. To increase the understanding of biochemical systems, a qualitative analysis method — especially a qualitative path analysis method — is introduced and applied to a case study. The applicability of low-level Petri nets for a qualitative analysis has been shown earlier. In this paper, a new complex hierarchical model of the glycolysis pathway and their interacting pathways is presented and analysed. The model is a generic one because it integrates the different pathways of various cell types into one unified view. At the end of this paper a meta-model illustrates the analysed invariants/ pathways.

Keywords: metabolic pathway, metabolic Petri net, P-invariant, T-invariant, glycolysis pathway

1 Introduction

Due to the rapidly growing amount of biologically experimental possibilities and the related amount of created experimental data, it is mandatory to transmit data in simple, analysable, and possibly validated model. Therefore, bioscientists need practicable, theoretically well-founded methods to construct, prove, analyse, and simulate a model, which is based on experimental data.

An aim of qualitative modelling is to predict the behaviour of a system only with information of the simple, atomic chemical reactions and their stoichiometric parameters. For some models, for example models based on differential equations, kinetic parameters are necessary to analyse a biochemical system. A given restriction of these models is often the imperfect and imprecise knowledge of these kinetic parameters, because up to now it is difficult to observe the processes in a cell at molecular level in vivo. Many known notations or models are insufficient for structural analysis or even animation of metabolic pathways.

Many different approaches for qualitative or quantitative analysis methods have been developed. For example a graph theoretical approach is described in [Ehrentreich03]. A mathematical approach is introduced in [Wiback02], computing a set of generating vectors that describe the conical steady-state solution space for flux distributions in a metabolic network, so called extreme pathways. But only Petri nets have been applied for both kinds of analysis. For example, quantitative Petri net models are introduced in [Chen03], [Genrich01], qualitative Petri net models are described in [Heiner01], [Koch04], and [Voss03].

In this paper, a concept will be introduced to construct a model, which can be used to determine metabolic pathways. This concept relies on Petri nets. Therefore, this concept needs only information of the simple stoichiometric chemical reactions and their stoichiometric parameters. All necessary information to determine every pathway in steady state is kept. This will be demonstrated by a non-trivial case study — the glycolysis pathway.

The paper is organized as follows. First, a biochemical case study — the glycolysis pathway — is introduced. All modelled pathways are shown with their total reaction equations. Some disadvantages of an often used notation are discussed. Secondly, some basic Petri net notions are recalled. Place-, Transition (P-/T-) invariants, and essential behavioural properties belong to the notions. A short chapter of biochemical notions follows. Especially, the differences between reaction types are recalled. The fifth chapter introduces the modelling concept. For each type of reactions an example is shown. A short discussion of the three possible modelling steps is given. Sixthly, a model of the pure glycolysis pathway is constructed and analysed. Existing P-/T-invariants of the model are validated and interpreted. In chapter seven the preliminary model is extended and analysed. The results are presented and interpreted on a new abstraction level. So, in this way a new meta-model was constructed. Last, some conclusions are given.

2 Case Study Glycolysis

The glycolysis is one of the main processes in human cells. In this paper we are going to construct a general unifying model for all human cells. It contains the glycolysis pathway (GP), the pentose-phosphate-pathway (PPP), the fructose-1-phosphate-pathway (F1PP), the fructose-6-phosphate-pathway (F6PP), and the galactose-glucose interconversion pathway (GGIP). Fructose (Fruc), galactose (Galac), and their pathways interact with the glycolysis. All the individual reactions take place in the cytoplasma of a cell. Glucose-6-phosphate or fructose-6-phosphate are the intermediate products of all described pathways. The described pathways start always with glucose (Gluc), fructose (Fruc) or galactose (Galac). Reaction products are lactate (Lac), pyruvate (Pyr), Ribose-5-phosphate (R5P), or/and NADPH. The gluconeogenesis — the inversion of the glycolysis— is not modelled, because some reactions of them do not take place in the cytoplasma of a cell. Moreover, the gluconeogenesis is only active in liver cells. An overview of the considered reactions can be found at the end of this paper.

2.1 Total reaction equations of each considered pathway

- **glycolysis**

aerobic: $\text{Glucose} + 2\ P_i + 2\ \text{ADP} + 2\ \text{NAD}^+ \rightarrow 2\ \text{Pyruvate} + 2\ \text{ATP} + 2\ \text{NADH} + 2\ H^+ + 2\ H_2O$

anaerobic $\text{Glucose} + 2\ P_i + 2\ \text{ADP} \rightarrow 2\ \text{Lactate} + 2\ \text{ATP} + 2\ H_2O$

- **fructose-1-phosphate -pathway** (only in cells of the liver)

  Fructose (Fruc) +2 ATP → Glyceraldehyde-3-phosphate (*GAP*) + Dihydroxyacetone phosphate (*DHAP*) + 2ADP + $2H^+$

- **fructose-6-phosphate -pathway**

  Fructose (Fruc) + ATP → Fructose-6-phosphate (F6P) + ADP + H^+

- **pentose-phosphate-pathway**

  Glucose-6-phosphate (*G6P*) + 2 NADP + H_2O → 2 NADPH + 2 H^+ + CO_2 + Ribose-5-phosphate (*R5P*)

  Glyceraldehyde-3-phosphate (*GAP*) + 2 Fructose-6-phosphate (*F6P*) → 3 Ribose-5-phosphate (*R5P*)

  3 Glucose-6-phosphate (*G6P*) + 6 NADP + 3 H_2O → Glyceraldehyde-3-phosphate (*GAP*) + 6 NADPH + 6 H^+ + 3 CO_2 + 2 Fructose-6-phosphate (*F6P*)

- **galactose-glucose interconversion pathway**

  Galactose (Galac) + ATP → Glucose-6-phosphate (G6P) + ADP + H^+

2.2 Inspection of an existing notation

Figure 2.1 shows a notation, which is often used by bioscientists. It demonstrates the glycolysis and the pentose-phosphate-pathways of an erythrocyte cell.

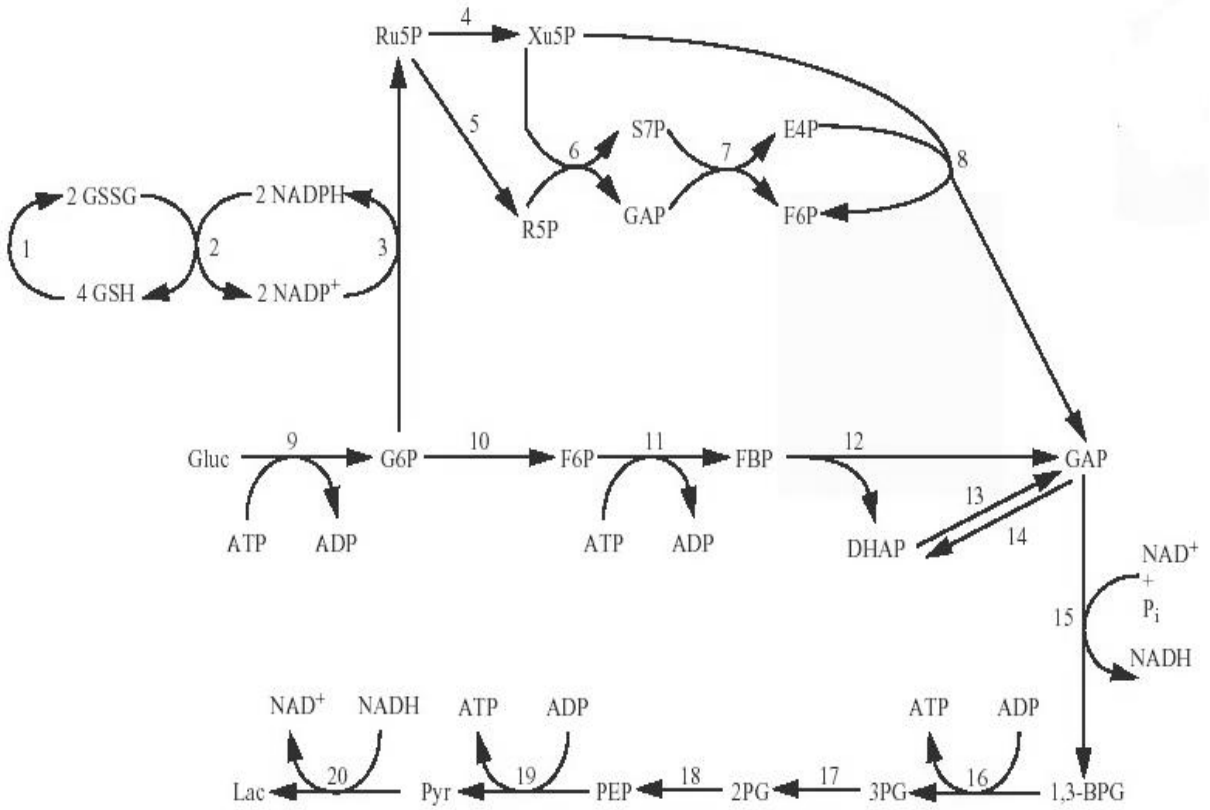


figure 2.1 example glycolysis and pentose-phosphate-pathway in the erythrocyte cell [Reddy96]

The following facts are worth mentioning.

- There are no differentiation between reaction types. For example, reaction 13 and 14 belong together. They form together an equilibrium reaction with a concentration of 96% on the side of DHAP. An analysis of this model can be incorrect.
- No kinetic information can be expressed by this notation.

- Furthermore, the model of the shown notation can not be directly analysed, simulated, and verified.
- In addition, GAP appears twice in the model. It is untreated, whether GAP exists at different locations or it has different states.

These disadvantages lead to the development of an improved model, which should be used instead to describe the biochemical system.

Reddy et. al. has shown in several papers that Petri nets can be used to model and analyse biochemical systems [Reddy93], [Reddy94], [Reddy96]. The model in Figure 2.1 served as starting point by Reddy and for the papers by Heiner et al. [Heiner98, Heiner01, Heiner02]. The ideas developed there form the basis for this paper, in which we are going to develop a generalised hierarchical model of the glycolysis and their essential pathways. Our model summarises all possible pathways of an arbitrary human cell, which supply or remove the intermediate products (G6P, F6P, GAP, and DHAP) of the glycolysis.

3 Basic Petri Net Notions

First, some Petri net notions will be recalled that we apply to glycolysis and the other modelled pathways later on. This chapter introduces Petri nets, their essential properties and analysis possibilities. For more detailed information [Starke90], [Baumgarten96], and [PNWorld] are recommended.

Let $N = (P, T, F, V, m_0)$ be a (low-level) Petri net with a set of places (P) and a set of transitions (T).

- $P = \{ p_1, p_2, .., p_m\}$ as finite set of places, $m \in \mathbb{N}^+$
- $T = \{ t_1, t_2, .., t_n\}$ as finite set of transitions, $n \in \mathbb{N}^+$
- $F \subseteq (P \times T) \cup (T \times P)$ as set of directed arcs between places and transitions or vice versa.
- $P \cap T = \emptyset$ and $P \cup T \neq \emptyset$
- V is a function, which assigns to every arc ((p, t) or (t, p)) a positive multiplicity ($V(f) \in \mathbb{N}^+$). The default is equal to one, if the arc is not labelled.
- m_0 is the initial marking

A marking M of a Petri net is a vector y of dimension m. Each component of a vector y contains the amount of tokens of a place. A marking represents a state of the modelled system.

Figure 3.1 depicts the graphical representation of all Petri net components. Three types of graphical representations, which mean all the same from an analysis point of view, exist for a transition. To avoid crossing arcs in a larger Petri net, logical places and transitions are established. Logical nodes with the same name and type represent the same place/transition.

Each transition is associated with a finite number of input places, also called pre-places (Ft), and output places, also called post-places (tF). A given transition is enabled, if the amount of tokens on all its pre-places is greater than or equal to the weights on the arcs connecting the places and this transition. One transition occurrence (firing) leads to a change of marking. A firing transition takes tokens from the pre-places (corresponding to the arc weights) and deposits tokens to its post-places (corresponding to the arcs weight). It is important to mention that no transition <u>must</u> fire.

This type of a Petri net, including the firing rule given, is called a Place-Transition net or shortly P/T net, and is classified as a low-level Petri net. Many extensions to P/T nets have been developed for various modelling and simulation purposes. These nets are often called high-level Petri nets. Some examples for using a high-level Petri net in systems biology are shown in [Voss03], [Heiner01], and [Genrich01] using coloured Petri nets, [Chen03] using

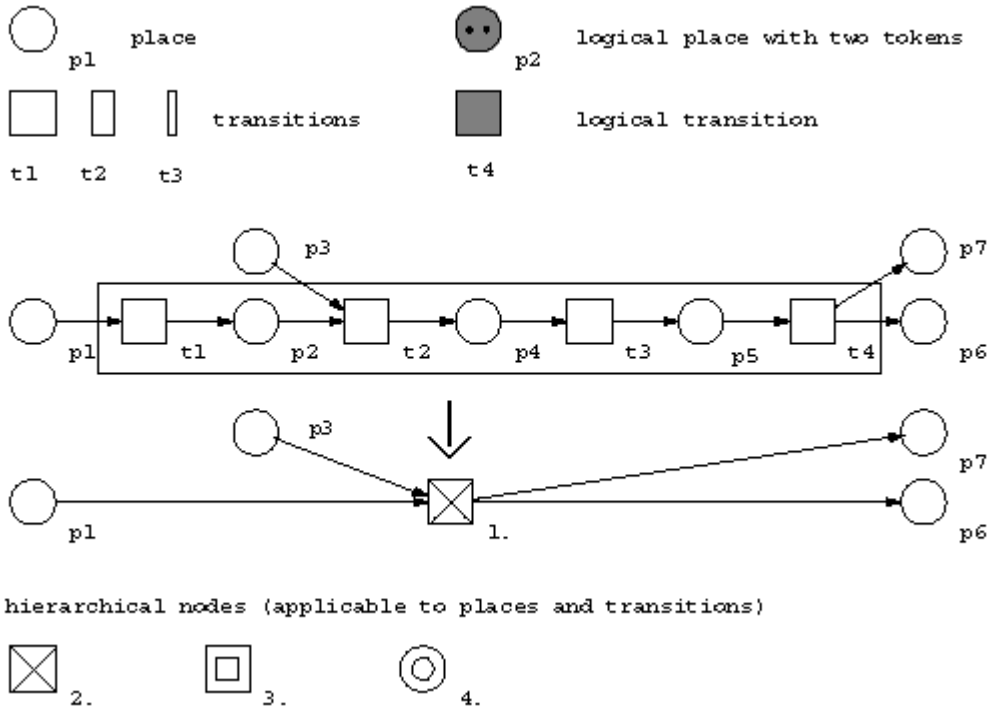


figure 3.1 typical components of a Petri net

hybrid Petri nets and [Matsuno03] using hybrid functional Petri nets. Another graphical component to promote clarity is the hierarchy concept. A hierarchical node increases the clarity without changing the semantics of the net. A Petri net class, which contains the hierarchy concept as an extension is also sometimes classified as a high-level Petri net. Our P/T net type contains the hierarchy concept, but for this paper we assume that this Petri net type is classified as a low-level Petri net.

3.1 Structural properties

Structural properties stand for all properties of a Petri net, which are independent of the initial marking. Transition (T-)invariants and Place (P-)invariants are essential to structural properties.

Incidence matrix

The (*m x n)* matrix *C* is called *incidence matrix* and is defined as follows.

i. $C=(c_{i,j})_{1\le i\le m,1\le j\le n}$ *with* $c_{i,j}:=\Delta t_j(p_i)$

ii. $\Delta t_j(p_i):=t_j^+-t_j^-$

iii. $t_j^+:=\begin{cases}V(t,p), & \text{if } p\in tF\\ 0, & \text{else}\end{cases}$ *iv.* $t_j^-:=\begin{cases}V(p,t), & \text{if } p\in Ft\\ 0, & \text{else}\end{cases}$

V(x,y) represents an arc weight of a place/transition or transition/place pair. In other words the elements of the incidence matrix $c_{i,j}$ are the positive/negative labels of the arcs pointing from/to transition t_j to/from place p_i. An example incidence matrix is shown in figure 3.3.

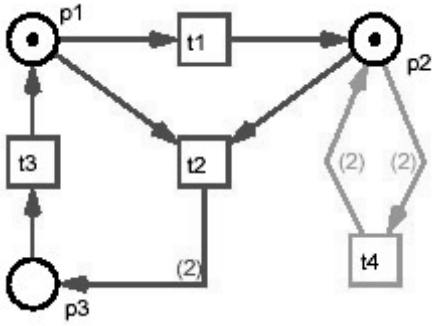


figure 3.2 Petri net with P-/T-invariants

$$\mathbf{C} = \begin{array}{c} P \backslash T \\ p_1 \\ p_2 \\ p_3 \end{array} \begin{array}{c} \begin{array}{cccc} t_1 & t_2 & t_3 & t_4 \end{array} \\ \begin{bmatrix} -1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 2 & -1 & 0 \end{bmatrix} \end{array}$$

figure 3.3 Incidence matrix of figure 3.2

T-invariant

Let a marking *M* be a state of a Petri net. A T-invariant describes a multi set of transitions, which restore this marking *M*. The solutions of the equation system

$$Cx = 0, \text{ with } x > 0 \text{ and } x \in \mathbb{N}$$

are called (*positive) T-invariant.* The T-invariant *x* is called *minimal*, if there exists no vector $y \geq 0$ with $supp(y) \subset supp(x)$, whereby supp(x) (read as support of x) describes the set of non-zero components in *x*, and the largest common divisor of all components of *x* is equal to one. The vectors *x*, *y* are of dimension *n.* Each component of them represents a transition. The set of minimal T-invariants forms a base. All T-invariants may be derived by non-negative integer linear combinations of the minimal T-invariants and by division with the largest common divisor. Each T-invariant represents a pathway through the Petri net. The Petri net is covered by T-invariants, if a T-invariant exists, which contains all transitions of the net.

Figure 3.2 contains all minimal T-invariants. The vectors $x_1 = (1,1,2,0)^T$ (dark grey) and $x2 = (0,0,0,1)^T$ (light grey) are these minimal T-invariants. The invariant x_1 can be interpreted as follows: Transition t_2 must fire once, t_3 twice, and t_1 once to reproduce the state m_0 = (1,1,0). The order of the transition firing is not determined by a T-invariant. Usually, this fact is not a problem, because the order is often not required, or the order can be derived from the model.

P-invariant

P-invariants are defined by solutions of the equation system

$$y^T C = 0, \text{ with } y > 0 \text{ and } y \in \mathbb{N}$$

The vectors *y* are called (*positive) P-invariants*. The number *z* of tokens of the invariant *y* is determined by the following formula.

$$y^T m_0 = z$$

m_0 is the initial state. The P-invariant *y* is called minimal, if there exists no vector $x \geq 0$ with $supp(x) \subset supp(y)$, and the largest common divisor of *y* is equal to one. The vectors *x*, *y* are of dimension *m.* Each component of them represents a place. The set of minimal P-invariants forms a base. All non-minimal P-invariants may be derived by non-negative integer linear

combinations of the minimal P-invariants and by division with the largest common divisor of a vector. A P-invariant of a biochemical net represents a substance preservation.

Only one P-invariant $y_1=(1,1,1)^T$ exists in the Petri net in the figure 3.2, covering the whole net. For this reason the number of tokens in the Petri net is always the same, here two.

It should be noticed that the computation of non-negative minimal P- or T-invariants, which form a unique generating system, leads back to the problem of solving systems of linear (in-) equations in non-negative integer variables. This problem is solvable by a non-deterministic Turing machine in polynomial-time, in other words the problem is *NP-complete* [Schrijver99].

3.2 Behavioural properties

In general, the behavioural properties depend on the initial marking of a Petri net. Because of this, it is necessary to provide additional information and analysis possibilities. There exist many analysable properties, which describe the modelled system behaviour. The following properties are the most essential ones for a Petri net.

liveness

A net is called live, if all transitions of the net are live at the initial marking. That means, there is no state reachable, where a transition is dead [Starke90]. A transition is dead, if it will be enabled never again.

boundedness

A Petri net is bounded, if all places of the net are bounded. A place is bounded, if the amount of tokens is limited at every marking reachable from the initial marking.

reversibility

A Petri net is called reversible, if the initial marking is reachable again from every reachable marking.

Liveness, boundedness, and reversibility are independent from each other. For this reason every combination of them is possible. So there exist eight possible combinations of the behavioural properties. The construction of a model aims often at a live and bounded Petri net. The analysis capabilities of an bounded net are not limited in contrast to unbounded nets.

The independence from the initial marking is one of the main advantages during the calculation of invariants of a Petri net. Due to this fact conclusions on the structure and the behaviour of the modelled system are possible.

In summary it can be said that the computation of P-/T-invariants is a structural analysis of the model. For the computation of T-invariants it is not necessary that the net is bounded. All possible pathways in a model can also be computed, if the net is unbounded. The computation of P-invariants is only possible for bounded places. The non-computability of the reachability graph of a net is another disadvantage of an unbounded net.

4 Biochemical Basic Notions

Here, some biochemical notions are recalled to avoid any misunderstanding. [Stryer02] serves as reference for biological and chemical questions.

Steady state

A steady state is a special system state, in which all internal substance concentrations are constant. "In a steady state, the concentrations of intermediates stay the same, even if the concentrations of starting materials and products are changing." [Stryer02, page 202]

ubiquitous compounds

The molecules like ADP, ATP, NAD+, Pi etc. are called ubiquitous compounds, because they are found in sufficiently large amounts in the cell [Voss03].

Direct reaction

In this paper, the classical, not reversible chemical reaction is called direct reaction.

example: Glucose (Gluc) + ATP → Glucose-6-phosphate (G6P) + ADP + H^+

Reversible reaction

Two reactions are hidden behind this term. Their are two complementary single reactions under the same conditions. The point is that only one of them is active over a larger time period. The direction and the rate of the reversible reaction are implicitly controlled by direct reactions. The direct reactions, and in correlation to them the thermodynamic environment, is the reason for the control of the direct reactions. [BioWeb][1] contains an animation, which illustrates the thermodynamic behaviour of the glycolysis and gluconeogenesis, both controlled by three direct reactions.

example: Glucose-6-phosphate (G6P) ↔ Fructose-6-phosphate (F6P)

Equilibrium reaction

Equilibrium reactions are similar to reversible reactions. The difference is that both reactions are active at the same time. The aim of an equilibrium reaction is a stable state, concerning to the directly involved compounds. Ideally, at this equilibrium state and under the same conditions the reaction is stopped. The amount of the involved compounds at the stable state does not need to be the same. The following example illustrates this.

example: Dihydroxyacetone phosphate (DHAP) (96%) ↔ Glyceraldehyde-3-phosphate (GAP) (4%)

These three reaction types are important for a correct quantitative analysis of the model. So, the equilibrium and reversible reactions have different effects on the dynamic behaviour of the model. The behaviour of a reversible reaction is mostly similar to the behaviour of a direct reaction, in contrast to the behaviour of an equilibrium reaction. But this is not important for a pure pathway analysis, which is based only on the structure of the model. A modelled reversible or equilibrium reaction results in a trivial T-invariant.

5 Modelling Chemical Reactions

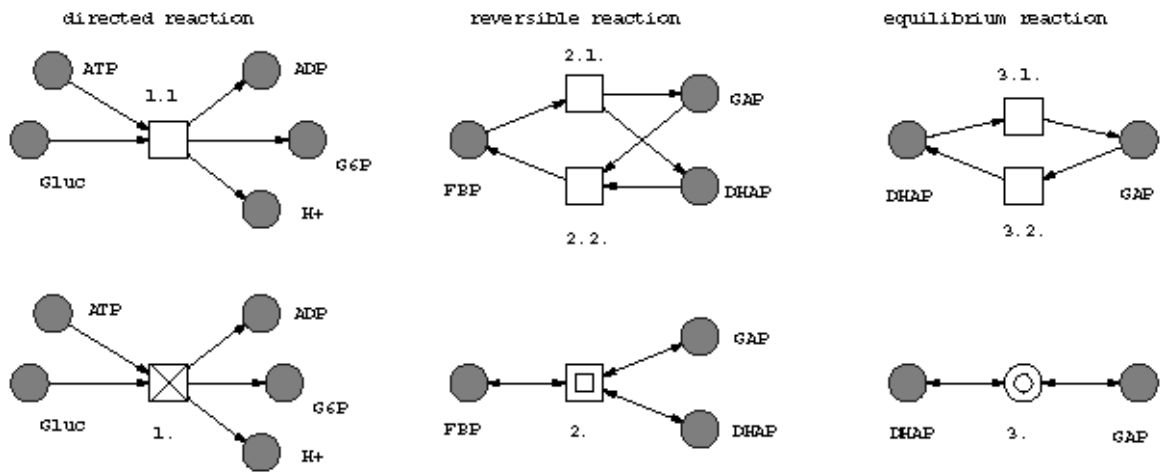
This chapter explains, how chemical reactions can be modelled with low-level Petri nets. The construction of a Petri net model for metabolic pathways is straightforward. Metabolites are modelled as places, reactions as transitions, and stoichiometric relations as directed arcs between them. Reddy et al. and Heiner et al. have applied these simple transformation rules on biochemical systems to get low-level Petri nets [Reddy93], [Reddy94], [Reddy96], [Heiner02].

1 See: "Conceptual Insights" → "Chapter 16" → "Energetics of Glucose Metabolism"

It is important that a given place represents only one metabolite on the same location. A transport from one location to another location has to be modelled as an additional transition.

Figure 5.1 shows the direct reaction, which is modelled as a single transition, the reversible reaction and the equilibrium reaction, which are modelled by two transitions (one for each direction). The differences between the latter ones is not modelled in the figure. For a pathway analysis this fact is not important. If other properties (especially behavioural properties) have to be analysed, an additional control logic for the reversible reactions is mandatory. This control logic has to represent the thermodynamic environment behaviour of the modelled system or in other word only one direction of the reversible reaction may be active at the same time.

For a better layout, all reactions are modelled using hierarchical nodes. Three hierarchical node types are used to distinguish the three reaction types. The crossed quadrate represents a directed reaction, the double quadrate a reversible reaction, and the double circle an equilibrium reaction. The structure of a given reaction is defined by its reaction type, contained in a hierarchical node. Furthermore, it should be noted that the double arrows stand for two arcs on top of each other. This is done to improve the readability. According to the type of reactions each single reaction is to be read from top to bottom and vice versa, or from left to right and vice versa. Reactions from left to right or top to bottom have the suffix ".1". All other reactions have the suffix ".2". For example, a single reaction from left to right is to understand as a reaction, where all products are on the right side of the transition and the starting materials on the left side.



1. Glucose (Gluc) + ATP → Glucose-6-phosphate (G6P) + ADP + H^+
2. Fructose biphosphate (FBP) ↔ Glyceraldehyde-3-phosphate(GAP) +Dihydroxyacetone-phosphate(DHAP)
3. Dihydroxyacetone-phosphate(DHAP) ↔ Glyceraldehyde-3-phosphate (GAP)

figure 5.1 example reactions and their hierarchical nodes for each reaction type (without the catalysing enzymes)

The models, presented in the following, are assumed to be in a steady state and the necessary amount of enzymes exists for the firing of a transition. For this reason, enzymes are not modelled, explicitly. Therefore, it is not possible to distinguish between a reversible and an equilibrium reaction without additional control logic. This is not important for the intended pathway analysis, because no possible pathway is removed. A model with additional control logic — using a high-level Petri net — is presented in [Heiner01] and [Voss03].

The construction of a basic model from the basic reaction equations is straightforward and simple. The figure 5.2 shows a model of a single reaction. The basic model is not sufficient for a pathway analysis. An environment of the modelled reactions has to be added. An exact description of the environment behaviour in relation to the real world is not practicable. But, the supply — modelled as a transition without pre-places — and the removal — modelled as a transition without post-places — as basic environment behaviour is necessary. The resulting model is shown in figure 5.2, which is bordered by transitions. Supply and removal transitions are drawn as small rectangle to improve the readability.

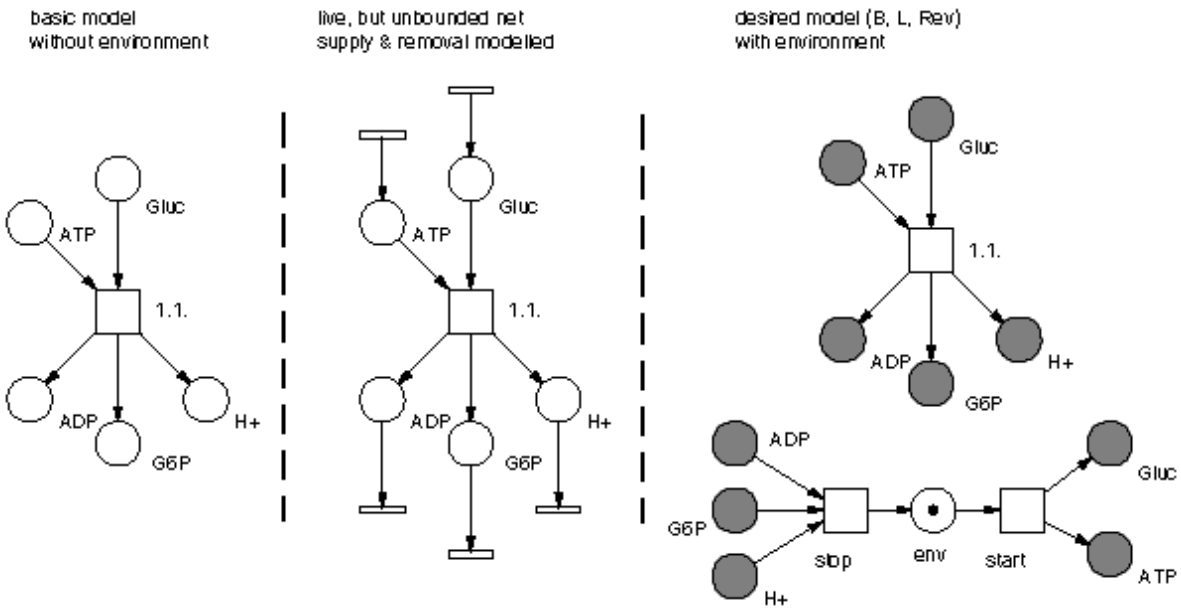


figure 5.2 general modelling steps

A model, including the environment, which is live, bounded, and reversible, is desirable, because its analysis and simulation is much easier. Because of the large scale of the basic model, it is difficult or hardly to ensure these properties by a manually constructed low-level Petri net. As mentioned in chapter 3, for a pathway analysis it is not necessary that the model is bounded. Information and some properties of a preliminary model can be reused in an elaborated model, for example, some P-/T-invariants or parts of them.

6 P-/T-Invariant Analysis of the Pure Glycolysis Model

In this chapter a model of the pure glycolysis (without interacting pathways) is constructed and analysed. It is used to introduce the glycolysis and some analysing possibilities of a live, bounded, and reversible Petri net. For the pure glycolysis it was possible to design an environment behaviour, so that the resulting Petri net has the desired properties.

6.1 Simplifications

Because the gluconeogenesis is not modelled, the thermodynamic benefit is on the side of the production of pyruvate or lactate. Most of the reversible reactions are surrounded by directed reactions. It is possible to transform a reversible reaction to a directed reaction by removing the following transitions 2.2, 4.2, 6.2, 7.2, 8.2, and 9.2.

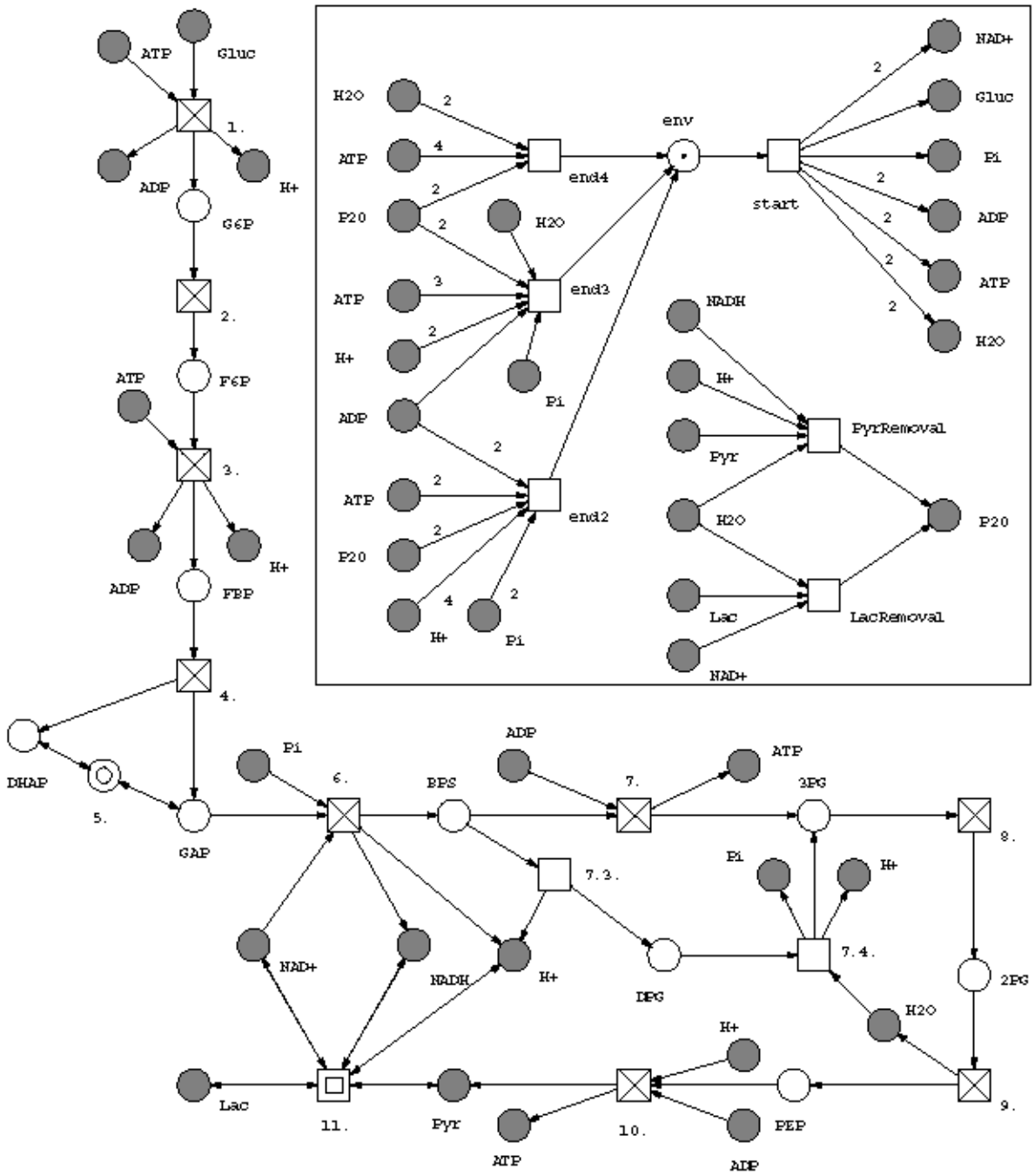


figure 6.1 pure glycolysis model

6.2 Comments, Compare figure 6.1

- The reactions 7.3 and 7.4 are a special part of the pathway. These reactions are only possible in the erythrocyte cells. For more information see [Stryer94, page 384] and [Kreutzig00, page 166]
- The introduced hierarchical nodes are applied to this model.

- The model can only be used for a qualitative analysis of the structure of the system, because a regulation through enzymes is not modelled. Furthermore, the modelled environment must not necessarily represent the real behaviour of the environment of the glycolysis. It is used to obtain the desired net properties.

6.3 Properties

The presented Petri net model is live, bounded, and reversible. Furthermore, it is covered by P- and T- invariants. Especially, a covering by T-invariants is important. With this information all modelled reactions contribute to some pathway. The following vectors are minimal T-invariants. Each component of the upper line is a transition name, each component of the middle line represents the corresponding enzyme name, if one exists, and each component of the lower line describes, how often the corresponding transition must fire to reproduce the initial state. The equations of the glycolysis (anaerobic, aerobic), introduced in chapter 2.1, correspond to the following two T-invariants.

$$\boldsymbol{t_1}=\begin{pmatrix} start & 1.1. & 2.1. & 3.1. & 4.1 & 5.1 & 6.1 & 7.1 & 8.1 & 9.1 & 10.1 & 11.1 & LacRemoval & end4 \\ start & hk. & pgi. & pfk. & al & tpi & gapA & pgk & gpm & eno & pyk & ldh & LacRemoval & end4 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \end{pmatrix}$$

$$\boldsymbol{t_2}=\begin{pmatrix} start & 1.1. & 2.1. & 3.1. & 4.1 & 5.1 & 6.1 & 7.1 & 8.1 & 9.1 & 10.1 & PyrRemoval & end4 \\ start & hk. & pgi. & pfk. & al & tpi & gapA & pgk & gpm & eno & pyk & PyrRemoval & end4 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \end{pmatrix}$$

From these invariants it can be derived that starting from one molecule of Glucose (Gluc) with the necessary substances (2 ADP, 2 P_i, and maybe NAD^+) either two lactate (Lac) or two pyruvate (Pyr) are created. It is also possible that one lactate and one pyruvate are created on account of reaction 4. This confirms the T-invariant t_3.

$$\boldsymbol{t_3}=\begin{pmatrix} start & 1.1. & 2.1. & 3.1. & 4.1 & 5.1 & 6.1 & 7.1 & 7.3. & 7.4. & 8.1 & 9.1 & 10.1 & 11.1 & LacRemoval & end3 \\ start & hk. & pgi. & pfk. & al & tpi & gapA & pgk & bpgm & bpgp & gpm & eno & pyk & ldh & LacRemoval & end3 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 1 \end{pmatrix}$$

$$\boldsymbol{t_4}=\begin{pmatrix} start & 1.1. & 2.1. & 3.1. & 4.1 & 5.1 & 6.1 & 7.3. & 7.4. & 8.1 & 9.1 & 10.1 & 11.1 & LacRemoval & end2 \\ start & hk. & pgi. & pfk. & al & tpi & gapA & bpgm & bpgp & gpm & eno & pyk & ldh & LacRemoval & end2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \end{pmatrix}$$

$$\boldsymbol{t_5}=\begin{pmatrix} start & 1.1. & 2.1. & 3.1. & 4.1 & 5.1 & 6.1 & 7.1 & 7.3. & 7.4. & 8.1 & 9.1 & 10.1 & PyrRemoval & end3 \\ start & hk. & pgi. & pfk. & al & tpi & gapA & pgk & bpgm & bpgp & gpm & eno & pyk & PyrRemoval & end3 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 1 \end{pmatrix}$$

$$\boldsymbol{t_6}=\begin{pmatrix} start & 1.1. & 2.1. & 3.1. & 4.1 & 5.1 & 6.1 & 7.3. & 7.4. & 8.1 & 9.1 & 10.1 & PyrRemoval & end2 \\ start & hk. & pgi. & pfk. & al & tpi & gapA & bpgm & bpgp & gpm & eno & pyk & PyrRemoval & end2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \end{pmatrix}$$

The T-invariants t_3, t_4, t_5, t_6 are <u>only</u> possible in an erythrocyte cell, because the reactions 7.3 and 7.4 produce less ATP.

The transitions (5.1, 5.2) and (11.1, 11.2) form also T-invariants, which are called trivial ones, because they represent an equilibrium reaction or a reversible reaction, respectively.

In general, two Glyceraldehyde-3-phosphate (GAP) are produced from one Glucose. From every produced Glyceraldehyde-3-phosphate (GAP) lactate or pyruvate is generated, depending on the presence or absence of oxygen (O_2). In summary two ATP are produced from one Glucose. An exception is the erythrocyte cell. In this cell, it is possible that no ATP at all is generated from Glucose. If pyruvate is generated, a NADH has to be consumed and re-oxidized.

The model is bounded. Therefore, P-invariants exist. In general, P-invariants represent a possible appearance of a given elementary molecule or a conservation of an elementary molecule. The places of NAD^+, NADH, “P20”, and “env” form a P-invariant. The

biochemical reason is that NAD^+ and NADH can be transformed into each other. The P-invariant contains the places “P20” and “env” only on account of the modelled environment.

$$\boldsymbol{p_1}=\begin{pmatrix} env & ATP & P_i & G6P & F6P & FBP & DHAP & GAP & BPS & DPG & 3PG & 2PG & PEP \\ 4 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 2 & 2 & 1 & 1 & 1 \end{pmatrix}$$

$$\boldsymbol{p_2}=\begin{pmatrix} env & Gluc & G6P & F6P & FBP & DHAP & GAP & BPS & DPG & 3PG & 2PG & PEP & Pyr & Lac & P20 \\ 2 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The shown P-invariants can be biochemically interpreted as follows. The vector p_1 represents the conservation of each phosphate atom. It is not necessary to compute the amount of tokens contained in a P-invariant, because, if place “env” contains one token, no other tokens are in the net. So “env” has a valency of four. If the transition “start” fires, then only ATP and P_i are affected. ATP and P_i have a valency of one. For this reason four phosphate atoms, each represented by one token, are contained in the net. The P-invariant p_2 corresponds to the conservation of each carbon atom. Each occurrence of the number one represents a group of three carbon atoms. For example the molecule Glucose contains six carbon atoms and the corresponding place has a valency of two. The place GAP has a valency of one and the molecule GAP contains three carbon atoms.

In special cases, there exist P-invariants, which can not be biochemically interpreted. For example the following one.

$$\boldsymbol{p_3}=\begin{pmatrix} env & NAD+ & BPS & DPG & 3PG & 2PG & PEP & Pyr & P20 \\ 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The invariant p_3 represents a pure net structure, but has no biochemical interpretation. A reason for this is a simple Petri net property, the undistinguishability of tokens in a low-level Petri net. With additional knowledge and a coloured Petri net model it should be possible to avoid such kind of P-invariants.

Another fact should be noted. The computed T-invariants depend on the modelled environment. For example the T-invariant t_1 without environment transitions (start, PyrRemoval, LacRemoval and end4) has a greater support (supp(t_1)=11) than the T-invariant t_2 without environment (supp(t_2)=10). Because of this t_1 would be no minimal T-invariant. This fact gets a greater significance in the next chapter, which contains an unbounded model.

Here, in contrast to [Reddy96] and [Heiner02] a more general model of the glycolysis pathway has been constructed. The model represents the pure glycolysis in all human cells. Different reaction types are introduced to refine the relations between the biochemical system and the Petri net model. The hydrogen ion (H^+) is modelled through an additional place separated from NADH, because it influences the rate of some reactions.

Some results of this model, especially the T-invariants, should also be valid for the next extended model.

7 Pathway Analysis of the Complex Glycolysis Model

Now, we are going to construct a more complex model consisting of the glycolysis pathway and the interacting pathways: the pentose-phosphate pathway, the fructose-1-phosphate pathway, the fructose-6-phosphate pathway, and the galactose-glucose interconversion pathway.

The previously modelled environment is deleted, because it is not valid anymore for the new model and replaced by a new one. The flat rectangles simulate the new environment behaviour, i.e. the supply and the removal of molecules. All grey places are logical places, which are used to improve the readability. Therefore, the model in figure 7.1 serves only for a

structural pathway analysis. The Petri net is live and unbounded. Therefore, no P-invariants exist. Because of the complexity of the model, and the later described possible pathways, it is impossible or really hard to construct by hand an environment without additional places and/or transitions, which control the path of one or more tokens. To solve this problem a high-level Petri net, for example a coloured Petri net, should be used. [Voss03] et. al. used a coloured Petri net to make the model bounded.

7.1 Simplifications, comments

The simplifications of the previous chapter are also valid for this model. Additionally, reactions 13.2 and 17.2 are removed, because they are also surrounded by directed reactions.

In the model, Reaction 26 needs an additional attention, because it ignores the reading rules described in chapter 5. The following formula describes reaction 26.

$$E4P + Xu5P \leftrightarrow F6P + GAP$$

A special view should be taken to the environment transitions (all transitions without pre-places or without post-places). These transitions describe only an arbitrary supply and removal of substances. But the expected minimal T-invariants depend on these transitions. Therefore, an simplified model without any *ubiquitous compounds* was constructed to obtain the “main” T-invariants. “main” T-invariant means a T-invariant without the influence of any ubiquitous compounds. The resulting T-invariants, described below, corresponds to the elementary modes computed by METATOOL, introduced in [Pfeiffer99]. The T-invariants correspond to elementary modes in METATOOL. Following places are removed: ATP, ADP, NAD^+, NADH, H^+, P_i and H_2O. An analysis of this complete model results in 32 “main” T-invariants and 8 trivial T-invariants. These T-invariants form a generating system computed by the tool “INA” [Ina03].

7.2 Properties

A T-invariant represents a single pathway or a combination of pathways. A detailed analysis of them follows.

The constructed model can be divided into two parts, part one and part two. The place GAP serves as main connection point between the two parts. This fact was realized during the construction of the model. For this reason we observe at first each part of the model. Later on, we conclude from the partial properties to the complete properties of the model. Part one consists of all transitions between the starting compounds (Gluc, Fruc, and Galac) and the intermediate product GAP including the pentose-phosphate pathway. Part two consists of all transitions between GAP and the products (Pyr and Lac). This partitioning simplifies the analysis. Part two is not changed in comparison to the model of the pure glycolysis. For this reason only part one has to be analysed in more detail. That's why we start with the simpler case. The supply and removal of GAP is omitted to simplify the later concatenation of the partial invariants. We start with the easier and smaller part.

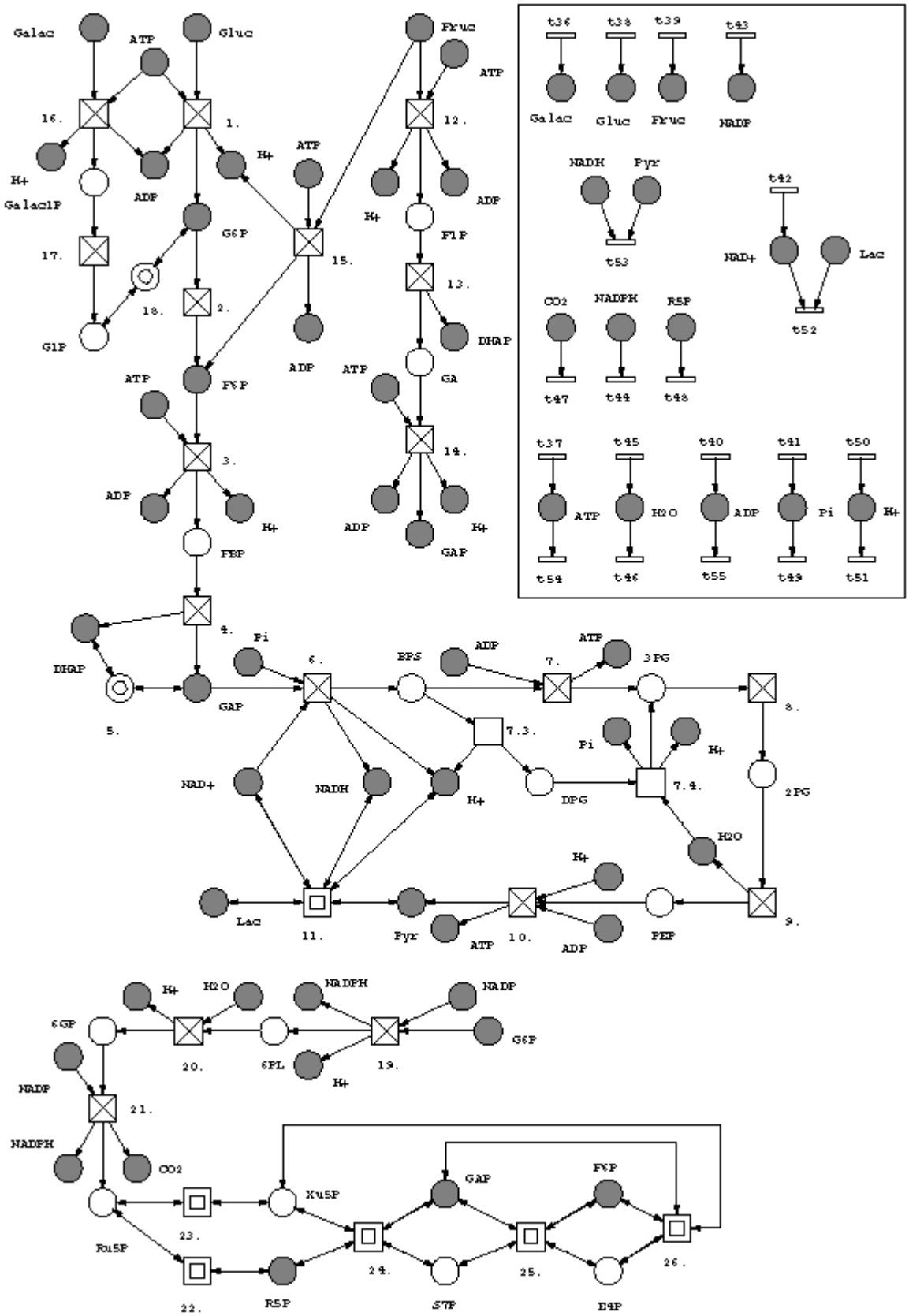


figure 7.1 complete structural glycolysis model with ubiquitous compounds

Part two: There exist four single paths through part two. One Pyr or Lac is generated from one GAP. Normally, in all human cells two ATP are regenerated from ADP for each GAP, except in the erythrocyte cell. In the erythrocyte cell it is possible that only one ATP is regenerated for each GAP. The following total reaction equations describe the four paths, and these reactions can be confirmed by their corresponding part of complete T-invariants. All following partial T-invariants contain only the given part of the model. The total reaction equations are not simplified. All necessary and generated substances are given in the equation.

Both partial paths *a*, *b* are possible in all human cells.

a) $GAP + 2\ ADP + P_i + NAD^+ + H^+ \rightarrow 2\ ATP + H_2O + Pyr + NADH + H^+$

b) $GAP + 2\ ADP + P_i + NAD^+ + H^+ \rightarrow 2\ ATP + H_2O + Lac + NAD^+$

In an erythrocyte cell the following two partial paths are additionally possible.

c) $GAP + ADP + P_i + NAD^+ + H_2O \rightarrow P_i + ATP + H_2O + Pyr + NADH + 2\ H^+$

d) $GAP + ADP + P_i + NAD^+ + H_2O \rightarrow P_i + ATP + H_2O + Lac + H^+$

t_a, t_b, t_c, and t_d are the corresponding T-invariants to the formulas *a*, *b*, *c,* and *d*.

$$\boldsymbol{t_a}=\begin{pmatrix} 6.1 & 7.1 & 8.1 & 9.1 & 10.1 & rePyr \\ gapA & pgk & gpm & eno & pyk & rePyr \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \qquad \boldsymbol{t_c}=\begin{pmatrix} 6.1 & 7.3 & 7.4 & 8.1 & 9.1 & 10.1 & rePyr \\ gapA & bpgm & bpgp & gpm & eno & pyk & rePyr \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\boldsymbol{t_b}=\begin{pmatrix} 6.1 & 7.1 & 8.1 & 9.1 & 10.1 & 11.1 & reLac \\ gapA & pgk & gpm & eno & pyk & ldh & reLac \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \qquad \boldsymbol{t_d}=\begin{pmatrix} 6.1 & 7.3 & 7.4 & 8.1 & 9.1 & 10.1 & 11.1 & reLac \\ gapA & bpgm & bpgp & gpm & eno & pyk & ldh & reLac \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Part one: The galactose is equivalent to Glucose contributing to all possible paths. This can be confirmed by the following formula and the discovered invariants. For this reason only Glucose is considered in the reaction equations.

$$Galac + ATP \rightarrow G6P + ADP + H^+$$
$$Gluc + ATP \rightarrow G6P + ADP + H^+$$

If a transition (Tn) fires, it is supposed that all involved enzymes are available. Four possible partial paths (t_{e1}, t_{e2}, t_f, and t_g) exist from Gluc, Galac, and Fructose to GAP without considering the pentose-phosphate pathway. These paths are active, if the cell needs energy in form of ATP and molecules of Pyr and/or Lac.

e) $Gluc + 2\ ATP \rightarrow 2\ GAP + 2\ H^+ + 2\ ADP$ (first part of glycolysis)

f) $Fruc + 2\ ATP \rightarrow 2\ GAP + 2\ H^+ + 2\ ADP$ (across Fructose-6-phosphate path, Tn 15)

g) $Fruc + 2\ ATP \rightarrow 2\ GAP + 2\ H^+ + 2\ ADP$ (across Fructose1phosphate path, Tn 12,13,14)

$$\boldsymbol{t_{e1}}=\begin{pmatrix} suGluc & 1.1 & 2.1 & 3.1 & 4.1 & 5.1 \\ suGluc & hk & pgi & pfk & al & tpi \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \qquad \boldsymbol{t_f}=\begin{pmatrix} suFruc & 15.1 & 3.1 & 4.1 & 5.1 \\ suFruc & hk & pfk & al & tpi \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\boldsymbol{t_{e2}}=\begin{pmatrix} suGalac & 16.1 & 17.1 & 18.1 & 2.1 & 3.1 & 4.1 & 5.1 \\ suGalac & galK & gal* & pgm & pgi & pfk & al & tpi \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \qquad \boldsymbol{t_g}=\begin{pmatrix} suFruc & 12.1 & 13.1 & 14.1 & 5.1 \\ suFruc & scrK & f1pa & tk & tpi \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The complete path *h* is active, if NADPH and R5P are needed. For this path, fructose is not possible as starting material. A combination with one of the paths *a*, *b*, *c* or *d* is not possible, because no GAP is generated. The invariant t_{h1} matches the "mode 2" in [Stryer02, page 569]. Mode 2 means that the needs for NADPH and ribose-5-phosphate (R5P) are balanced. In this context Galactose is not considered in Stryer.

h) $Gluc + ATP + 2\ NADP + H_2O \rightarrow R5P + 2\ NADPH + ADP + 3\ H^+ + CO_2$

$$t_{h1}=\begin{pmatrix} suGluc & 1.1 & 19.1 & 20.1 & 21.1 & 22.1 \\ suGluc & hk & g6pdh & 6\,pgl & 6\,pgd & rpi \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad t_{h2}=\begin{pmatrix} suGalac & 16.1 & 17.1 & 18.1 & 19.1 & 20.1 & 21.1 & 22.1 \\ suGalac & galK & gal* & pgm & g6pdh & 6\,pgl & 6\,pgd & rpi \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The invariant t_i matches the "mode 4" in [Stryer02, page 570]. Mode 4 means that both NADPH and ATP are required. Three molecules of Gluc are necessary to transform Ru5P in GAP. This mode also matches partially the "mode 3" in [Stryer02, page 569] except with a reaction of the gluconeogenesis. Mode 3 means that much more NADPH than ribose-5-phosphate (R5P) is required. Therefore, mode 3 is not completely contained in this model.

i) 3 Gluc + 5 ATP + 6 NADP + 3 H_2O → 6 NADPH + 5 GAP + 5 ADP +11 H^+ + 3 CO_2

$$t_{i1}=\begin{pmatrix} suGluc & 1.1 & 19.1 & 20.1 & 21.1 & 22.1 & 23.1 & 24.1 & 25.1 & 26.1 & 3.1 & 4.1 & 5.1 \\ suGluc & hk & g6pdh & 6\,pgl & 6\,pgd & rpi & rpe & tkt & tal & tkt & pfk & al & tpi \\ 3 & 3 & 3 & 3 & 3 & 1 & 2 & 1 & 1 & 1 & 2 & 2 & 2 \end{pmatrix}$$

$$t_{i2}=\begin{pmatrix} suGalac & 16. & 17.1 & 18.1 & 19.1 & 20.1 & 21.1 & 22.1 & 23.1 & 24.1 & 25.1 & 26.1 & 3.1 & 4.1 & 5.1 \\ suGalac & galK & gal* & pgm & g6pdh & 6\,pgl & 6\,pgd & rpi & rpe & tkt & tal & tkt & pfk & al & tpi \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 & 1 & 2 & 1 & 1 & 1 & 2 & 2 & 2 \end{pmatrix}$$

Some attractive invariants are the following. The invariants (t_{j1}, t_{j2}, t_k ,t_l, and t_m) represent paths, which are mainly active, if R5P and not too much NADPH are needed. It is also possible that NADP is not available. These four invariants are not minimal by the previously mentioned criteria. The support of the invariants t_{n1}, t_{n2}, t_o, t_p, and t_q are smaller than the support of these invariants. This is an interesting point, because through the production of GAP it can be combined with one of the invariants of part two. INA does not compute these invariants, if the model without the ubiquitous compounds is used.

j) 3 Gluc + 4 ATP → 3 R5P + GAP + 4 ADP + 4 H^+

k) 3 Fruc + 4 ATP → 3 R5P + GAP + 4 ADP + 4 H^+ (all three across Fructose-6-phosphate path)

l) Fruc + 2 Gluc + 4 ATP → 3 R5P + GAP + 4 ADP + 4 H^+ (Fruc across Fructose-1-phosphate path)

m) 3 Fruc + 4 ATP → 3 R5P + GAP + 4 ADP + 4 H^+ (one Fruc across Fructose-1-phosphate path and two across Fructose-6-phosphate path)

$$t_{j1}=\begin{pmatrix} suGluc & 1.1 & 2.1 & 3.1 & 4.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P & reGAP \\ suGluc & hk & pgi & pfk & al & tpi & tkt & tal & tkt & rpe & rpi & reR5P & reGAP \\ 3 & 3 & 3 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 1 \end{pmatrix}$$

$$t_{j2}=\begin{pmatrix} suGalac & 16. & 17.1 & 18.1 & 2.1 & 3.1 & 4.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P & reGAP \\ suGalac & galK & gal* & pgm & pgi & pfk & al & tpi & tkt & tal & tkt & rpe & rpi & reR5P & reGAP \\ 3 & 3 & 3 & 3 & 3 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 1 \end{pmatrix}$$

$$t_{k}=\begin{pmatrix} suFruc & 15.1 & 3.1 & 4.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P & reGAP \\ suFruc & hk & pfk & al & tpi & tkt & tal & tkt & rpe & rpi & reR5P & reGAP \\ 3 & 3 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 1 \end{pmatrix}$$

$$t_{l1}=\begin{pmatrix} suGluc & suFruc & 1.1 & 2.1 & 12.1 & 13.1 & 14.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P & reGAP \\ suGluc & suFruc & hk & pgi & scrK & f1pa & tk & tpi & tkt & tal & tkt & rpe & rpi & reR5P & reGAP \\ 2 & 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 1 \end{pmatrix}$$

$$t_{l2}=\begin{pmatrix} suGalac & suFruc & 16.1 & 17.1 & 18.1 & 2.1 & 12.1 & 13.1 & 14.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P & reGAP \\ suGalac & suFruc & galK & gal* & pgm & pgi & scrK & f1pa & tk & tpi & tkt & tal & tkt & rpe & rpi & reR5P & reGAP \\ 2 & 1 & 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 1 \end{pmatrix}$$

$$t_{m}=\begin{pmatrix} suFruc & 15.1 & 12.1 & 13.1 & 14.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P & reGAP \\ suFruc & hk & scrK & f1pa & tk & tpi & tkt & tal & tkt & rpe & rpi & reR5P & reGAP \\ 3 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 1 \end{pmatrix}$$

The formulas *n*, *o*, *p*, *q* are extensions of the paths *j*, *k*, *l,* and *m*. The additional molecules are mandatory to transform all generated GAP to R5P. NADPH or energy in form of ATP are not

produced. A combination of *n*, *o*, *p,* and *q* with *a*, *b*, *c,* and *d* is impossible, because GAP is completely transformed.

n) 5 Gluc + 6 ATP → 6 R5P + 6 ADP + 6 H^+

o) 5 Fruc + 6 ATP → 6 R5P + 6 ADP + 6 H^+ (all five across Fructose-6-phosphate path)

p) Fruc + 4 Gluc + 6 ATP → 6 R5P + 6 ADP + 6 H^+ (Fruc across Fructose-1-phosphate path)

q) 5 Fruc + 6 ATP → 6 R5P + 6 ADP + 6 H^+ (one Fruc across Fructose-1-phosphate path)

$$\boldsymbol{t_{n1}}=\begin{pmatrix} suGluc & 1.1 & 2.1 & 3.1 & 4.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P \\ suGluc & hk & pgi & pfk & al & tpi & tkt & tal & tkt & rpe & rpi & reR5P \\ 5 & 5 & 5 & 1 & 1 & 1 & 2 & 2 & 2 & 4 & 4 & 6 \end{pmatrix}$$

$$\boldsymbol{t_{n2}}=\begin{pmatrix} suGalac & 16.1 & 17.1 & 18.1 & 2.1 & 3.1 & 4.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P \\ suGalac & galK & gal* & pgm & pgi & pfk & al & tpi & tkt & tal & tkt & rpe & rpi & reR5P \\ 5 & 5 & 5 & 5 & 5 & 1 & 1 & 1 & 2 & 2 & 2 & 4 & 4 & 6 \end{pmatrix}$$

$$\boldsymbol{t_{o}}=\begin{pmatrix} suFruc & 15.1 & 3.1 & 4.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P \\ suFruc & hk & pfk & al & tpi & tkt & tal & tkt & rpe & rpi & reR5P \\ 5 & 5 & 1 & 1 & 1 & 2 & 2 & 2 & 4 & 4 & 6 \end{pmatrix}$$

$$\boldsymbol{t_{p1}}=\begin{pmatrix} suGluc & suFruc & 1.1 & 2.1 & 12.1 & 13.1 & 14.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P \\ suGluc & suFruc & hk & pgi & scrK & f1pa & tk & tpi & tkt & tal & tkt & rpe & rpi & reR5P \\ 4 & 1 & 4 & 4 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 4 & 4 & 6 \end{pmatrix}$$

$$\boldsymbol{t_{p2}}=\begin{pmatrix} suGalac & suFruc & 16.1 & 17.1 & 18.1 & 2.1 & 12.1 & 13.1 & 14.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P \\ suGalac & suFruc & galK & gal* & pgm & pgi & scrK & f1pa & tk & tpi & tkt & tal & tkt & rpe & rpi & reR5P \\ 4 & 1 & 4 & 4 & 4 & 4 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 4 & 4 & 6 \end{pmatrix}$$

$$\boldsymbol{t_{q}}=\begin{pmatrix} suFruc & 15.1 & 12.1 & 13.1 & 14.1 & 5.1 & 26.2 & 25.2 & 24.2 & 23.2 & 22.1 & reR5P \\ suFruc & hk & scrK & f1pa & tk & tpi & tkt & tal & tkt & rpe & rpi & reR5P \\ 5 & 4 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 4 & 4 & 6 \end{pmatrix}$$

The last eight T-invariants t_j, t_k, t_l, t_m, t_n, t_o, t_p, and t_q match the "mode 1" in [Stryer02, page 568]. Mode 1 means that much more ribose-5-phosphate than NADPH is required.

After analysing each part it is interesting, which paths can be combined and how many possibilities exist. Only the T-invariants *a, b, c, d, e, f, g, h, i, n, o, p* and *q* are considered. It is assumed that these invariants are minimal.

The following complete invariants are computed by INA.

$t_{e1} + 2t_a$	$t_{e1} + 2t_b$	$t_{e1} + 2t_c$	$t_{e1} + 2\ t_d$	t_{h1}	t_{h2}
$t_{e2} + 2t_a$	$t_{e2} + 2t_b$	$t_{e2} + 2t_c$	$t_{e2} + 2\ t_d$	t_{n1}	t_{n2}
$t_f + 2t_a$	$t_f + 2t_b$	$t_f + 2t_c$	$t_f + 2\ t_d$	t_o	
$t_g + 2t_a$	$t_g + 2t_b$	$t_g + 2t_c$	$t_g + 2\ t_d$	t_{p1}	t_{p2}
$t_{i1} + 5t_a$	$t_{i1} + 5t_b$	$t_{i1} + 5t_c$	$t_{i1} + 5\ t_d$	t_q	
$t_{i2} + 5t_a$	$t_{i2} + 5t_b$	$t_{i2} + 5t_c$	$t_{i2} + 5\ t_d$	→ 32 non-trivial possibilities	

It should be noticed that a general model of a human cell was constructed, so there exist some partial paths that can not be combined, because they are active in different types of a cell. The large amount of combinations confirms the complexity of the glycolysis and the interacting pathways. Figure 7.2 depicts all possible paths. The figure presents a new abstraction level of the Petri net model. Arcs of the same type and grey value belong together. The arcs with different types and grey values can mostly be combined. Stoichiometric parameters are not shown in figure 7.2. R5P is not directly generated from NADPH. It is only used to visualise the specific paths.

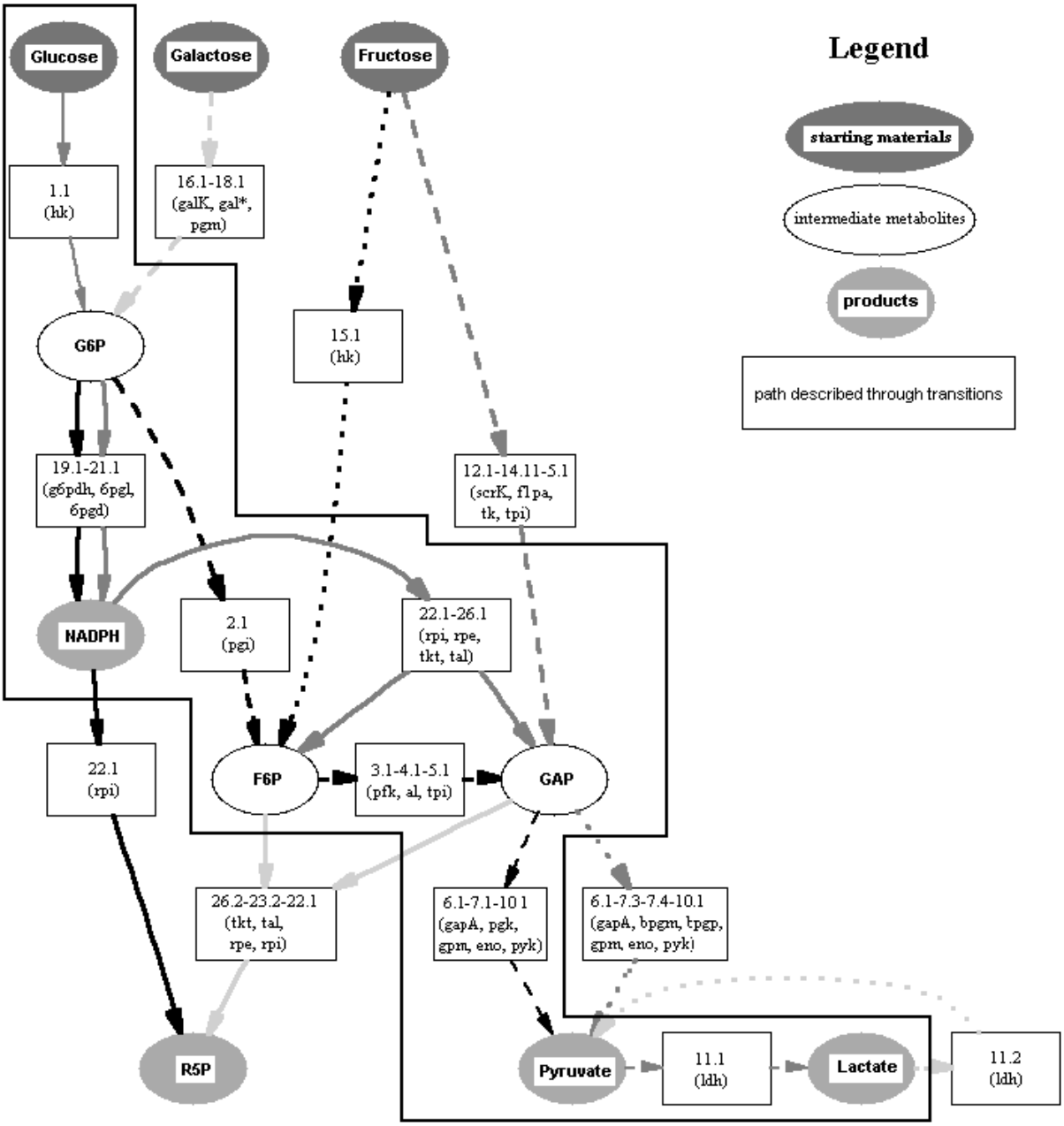


figure 7.2 schematic presentation of all possible paths in relation to the glycolysis pathway

In comparison to the models of [Reddy96] and [Heiner01], more pathways are modelled. Reddy's and Heiner's models are confined by the black polygon. The three modes (1, 2, and 4), according [Stryer02], are also included. Mode 3, according to [Stryer02], is not included, because a reaction of the gluconeogenesis would be necessary. If the gluconeogenesis would been modelled, even more combinations would arise. But, the gluconeogenesis and the glycolysis are reciprocally regulated. Therefore, only one direction of the reversible reactions is possible at the same time.

8 Conclusions

In this paper a generic low-level Petri net model of the glycolysis pathway and the pentose-phosphate-pathway for all human cells was constructed. This is a more complex model than the glycolysis pathway and the pentose-phosphate-pathway in an erythrocyte cell. The model illustrates the glycolysis pathway and all other pathways, which interact with the intermediate products of the glycolysis pathway. These pathways are the pentose-phosphate-pathway, the fructose-1-phosphate-pathway, the fructose-6-phosphate-pathway and the galactose-glucose interconversion pathway.

The model is analysed and validated. All existing T-invariants can be interpreted biochemically. There exist some P-invariants, which can not be biochemically interpreted. Such non-interpretable P-invariants may be avoidable by using another abstraction level and additional knowledge about the system. Especially the T-invariant analysis is a comfortable method to discover all possible interacting pathways and to validate the model.

In summary it can be said that it is quite easy to construct a low-level Petri net from a biochemical system, if the stoichiometric reaction equations are known. The construction is straightforward. Furthermore, the created Petri net can be validated by the computation of T-invariants. P-invariants are also suitable to validate the model, but here it is necessary that the model or the interesting parts of them are bounded. With increasing complexity of the biochemical system it gets more and more difficult or impossible to construct an environment behaviour by hand to make the net bounded. So it can be practicable to use a high-level Petri net instead, but the construction requires more sophisticated skills than the construction of a low-level Petri net. Moreover, there might be less analysis techniques available.

Acknowledgement

I like to thank Monika Heiner, Brandenburg University of Technology Cottbus, and Ina Koch, Technical University of Applied Sciences Berlin, for the fruitful discussions and hints.

References

[Baumgarten96] Baumgarten, B.: Petri-Netze Grundlagen und Anwendungen, Spektrum Akademischer Verlag, 1996

[BioWeb] Biochemistry, Fifth Edition, Additional Resources, http://bcs.whfreeman.com/biochem5/

[Chen03] Chen, M.; Hofestädt, R.: Quantitative Petri Net Model of Gene Regulated Metabolic Networks in the Cell; In Silico Biol, 3(3):347-365, 2003

[Ehrentreich03] Ehrentreich, F.; Schomburg, D.: Dynamic generation and qualitative analysis of metabolic pathways by a joint database/graph theoretical approach; Funct Integr Genomics, 3(4):189-196, 2003.

[Genrich01] Genrich, H.; Küffner, R.; Voss, K.: Executable Petri Net Models for the Analysis of Metabolic Pathways; International Journal on Software Tools for Technology (STTT), 3(4):394-404, 2001.

[Heiner98] Heiner, M: Application of Petri Nets on Metabolic Networks, Slides of the talk given at Gratuiertenkolleg "Dynamik und Evolution zellulärer und makromolekularer Prozesse", Prof. Heinrich, HU Berlin, department of biology, August 1998

[Heiner01] Heiner, M.; Koch, I.; Voss, K.: Analysis and Simulation of Steady States in Metabolic Pathways with Petri Nets, Proc. CPN Workshop, Univ. of Aarhus, August 2001, pp. 15 – 34

[Heiner02] Heiner, M.; Koch, I.: Biopathways and Petri Nets, Slides of the talk given in Berlin at Charite/Neuropädiatrie, Sept. 2002

[Ina03] INA – Integrated Net Analyzer v2.2, http://www.informatik.hu-berlin.de/~starke/ina.html

[Koch04] Koch, I.; Junker, B.H.; Heiner, M.: Application of Petri net theory to model validation of the sucrose breakdown pathway in the potato tuber, submitted to Bioinformatics, 2004.

[Kreutzig00] Kreutzig, T.: Kurzlehrbuch Biochemie, Urban & Fischer, 2000

[Löffler00] Löffler, G.: Basiswissen Biochemie mit Pathobiochemie, Springer, 2000

[Matsuno03] Matsuno, H.; Fujita, S.; Doi, A.; Nagasaki, M.; Miyano, S.: Towards Biopathway Modeling and Simulation. Proc. ICATPN 2003, LNCS 2679, pp.3-22

[Pfeiffer99] Pfeiffer, T.; Sanchez-Valdenebro, I.; Nuno,J.C.; Montero, F.; Schuster, S. METATOOL: for studying metabolic networks. Bioinformatics, 15(3):251-257, 1999

[PNWorld] Petri Nets World, http://www.daimi.au.dk/PetriNets/

[Reddy93] Reddy, V. N.; Mavrovouniotis, M. L.; Liebman, M. N: Petri Net Representation in Metabolic Pathways. Proc. First International Conference on Intelligent Systems for Molecular Biology, AAAI Press, Menlo Park, 1993, pp. 328-336

[Reddy94] Reddy, V.N.: Modeling Biological Pathways: A Discrete Event Systems Approach, Master's Degree, University of Maryland, 1994

[Reddy96] Reddy, V. N., Liebman, M. N. & Mavrovouniotis, M. L.: Qualitative Analysis of Biochemical Reaction Systems, Computers in Biology and Medicine 26 (9-24), 1996

[Schrijver99] Schrijver, A.: Theory of Linear and Integer Programming, John Wiley & Sons, 1999

[Starke90] Starke, P.H.: Analyse von Petri-Netz-Modellen, Teubner, 1990

[Stryer94] Stryer, L.: Biochemie, Spektrum Akademischer Verlag, 1994

[Stryer02] Stryer, L.: Biochemistry 5th Edition, W. H. Freeman, 2002

[Voss03] Voss, K.; Heiner, M.; Koch, I.: Steady State Analysis of Metabolic Pathways Using Petri Nets; *In Silico Biol.* **3**, 0031 (2003)

[Wiback02] Wiback, S.J; Palsson, B.O.; Extreme Pathway Analysis of Human Red Blood Cell Metabolism; Biophysical Journal, 83(2):808-818, 2002.

Abbreviations

Chemical reaction formulas, the modelling basis

GP	tk: GA + ATP → GAP + ADP + H
hk: Gluc + ATP → G6P + ADP + H	**F6PP**
pgi: G6P → F6P	hk2: Fruc + ATP → F6P + ADP + H
pfk: F6P + ATP → FBP + ADP + H	**GGIP**
al: FBP → DHAP + GAP	galK: Galac + ATP → Galac1P + ADP + H
tpi: DHAP ↔ GAP	gal: Galac1P → G1P
gapA: GAP + Pi + NAD → NADH + H + BPS	pgm: G1P ↔ G6P
pgk: BPS + ADP → PG3 + ATP	**PPP**
bpgm: BPS → DPG + H	g6pdh: G6P + NADP → 6PL + NADPH + H
bpgp: DPG + H2O → 3PG + Pi	6pgl: 6PL + H2O → 6GP + H
gpm: 3PG → 2PG	6pgd: 6GP + NADP → NADPH + CO2 + Ru5P
eno: 2PG → H2O + PEP	rpi: Ru5P ↔ R5P
pyk: PEP + ADP + H → ATP + Pyr	rpe: Ru5P ↔ Xu5P
ldh: Pyr + H + NADH ↔ Lac + NAD	tkt: Xu5P + R5P ↔ GAP + S7P
F1PP	tal: GAP + S7P ↔ F6P + E4P
scrK: Fruc + ATP → F1P + ADP + H	tkt2: Xu5P + E4P ↔ GAP + F6P
f1pa: F1P → DHAP + GA	

Metabolites / Compounds

2PG	2-Phosphoglycerate	GAP	Glyceraldehyde-3-phosphate
3PG	3-Phosphoglycerate	Gluc	Glucose
6GP	6-Phosphogluconate	Lac	Lactate
6PL	6-Phosphoglucono-δ-lactone	NAD^+	Nicotinamide adenine dinucleotide, oxidized form
ADP	Adenosine diphosphate		
ATP	Adenosine triphosphate	NADH	Nicotinamide adenine dinucleotide, reduced form
BPS	1,3-Biphosphoglycerate		
DHAP	Dihydroxyacetone phosphate	$NADP^+$	Nicotinamide adenine dinucleotide phosphate, oxidized form
DPG	2,3-Biphosphoglycerate		
E4P	Erythrose-4-phosphate	NADPH	Nicotinamide adenine dinucleotide phosphate, reduced form
F1P	Fructose-1-phosphate		
F6P	Fructose-6-phosphate	PEP	Phosphoenolpyruvate
FBP	Fructose-1,6-biphosphate	P_i	Orthophosphate, ionic form
Fruc	Fructose	Pyr	Pyruvate
G1P	Glucose-1-phosphate	R5P	Ribose-5-phosphate
G6P	Glucose-6-phosphate	Ru5P	Ribulose-5-phosphate
GA	Glyceraldehyde	S7P	Sedoheptulose-5-phosphate
Galac	Galactose	Xu5P	Xylulose-5-phosphate
Galac1P	Galactose-1-phosphate		

Correspondence between Petri net transition/reaction numbers, abbreviations, and enzymatic reactions

1.	hk	Hexokinase	15.	hk	Hexokinase
2.	pgi	Phosphoglucose isomerase	16.	galK	Galactokinase
3.	pfk	Phosphofructokinase	17.	gal*	Galactose 1-phosphate uridyl transferase, UDP-Galactose 4-epimerase
4.	al	Aldolase			
5.	tpi	Triose phosphate isomerase			
6.	gapA	GAP dehydrogenase	18.	pgm	Phosphoglucomutase
7.	pgk	Phosphoglycerate kinase	19.	g6pdh	
7.3.	bpgm	Bisphosphoglycerate mutase			Glucose 6-phosphate dehydrogenase
7.4.	bpgp	Bisphosphoglycerate phosphatase	20.	6pgl	Lactonase
8.	gpm	Phosphoglycerate mutase	21.	6pgd	6-Phosphogluconate dehydrogenase
9.	eno	Enolase	22.	rpi	Phosphopentose isomerase
10.	pyk	Pyruvate kinase	23.	rpe	Phosphopentose epimerase
11.	ldh	Lactate dehydrogenase	24.	tkt	Transketolase
12.	scrK	Fructokinase	25.	tal	Transaldolase
13.	f1pa	Fructose 1-phosphate aldolase	26.	tkt	Transketolase
14.	tk	Triose kinase			