Topological analysis of metabolic and regulatory networks by decomposition methods

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor rerum naturalium
(dr. rer. nat.)
im Fach Biophysik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von
Frau Dipl.-Inf. Ionela Oancea
geboren am 22.02.1977 in Bukarest

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Dr. Michael Linscheid

Gutachter:

1. Prof. Dr. Reinhart Heinrich
2. Prof. Dr. Hermann-Georg Holzhütter
3. Prof. Dr. Stefan Schuster

eingereicht am: 21. Juli 2003
Tag der mündlichen Prüfung: 1. Dezember 2003

Zusammenfassung

Die lebenden Organismen sind für eine wissenschaftliche Analyse zu kompliziert, wenn man sie als Ganzes und in ihrer vollen Komplexität betrachtet. Die vorliegende Arbeit behandelt die topologischen Eigenschaften von zwei wichtigen Teilen der lebenden Organismen: die metabolischen und die regulatorischen Systeme. Topolgische Eigenschaften sind solche, die durch die Netzwerkstruktur bedingt werden. Ein Signalsystem ist eine spezielle Art von regulatorischem System. Zwischen den metabolischen und Signalnetzen gibt es wichtige Unterschiede, die ihre Behandlung in unterschiedlicher Weise erfordert. In der metabolischen Pfadanalyse ist das Konzept der elementaren Flussmoden bereits als ein passendes Instrument für die Charakterisierung der einfachsten essentiellen Wege in biochemischen Systemen etabliert. Wir untersuchen die Eigenschaften und Vorteile dieses Konzepts in einigen besonderen Fällen.

Zuerst untersuchen wir die vielfach vorkommenden Enzyme mit niedriger Spezifität (z.B. Nukleosiddiphosphokinase, Uridinkinase, Transketolase, Transaldolase). Sie können parallel verschiedene Substrate und Produkte umwandeln. Auch die Enzym-Mechanismen sind vielfältig, wie wir mit dem Reaktionsschema für bifunktionelle Enzyme veranschaulichen. Wir betrachten dabei nur den Fall, dass ein bestimmtes aktives Zentrum mehrere Reaktionen katalysiert. Der Fall, dass das studierte Enzym mehrere solche aktiven Zentren hat, kann in den Fall mehrerer Enzyme transformiert werden, die nur ein aktives Zentrum haben. Wenn eine Krankheit das Ausgangsenzym ändert, werden dann in der Analyse auch alle ersetzenden Enzyme geändert. Es gibt zwei unterschiedliche Betrachtungsweisen, um multifunktionelle Enzyme zu beschreiben. Zum einen kann man die Gesamtreaktionen betrachten und zum anderen die elementaren Reaktionsschritte (Hemireaktionen, Halbreaktionen). Für Enzyme mit zwei oder mehr Funktionen ist es wichtig, nur linear unabhängige Funktionen zu betrachten, weil sonst zyklische elementare Moden auftreten würden, die keine Nettoumwandlung durchführen. Jedoch ist die Wahl der linear unabhängigen Funktionen nicht *a priori* eindeutig. Wir stellen eine Methode für das Treffen dieser Wahl vor, indem wir die konvexe Basis des Hemireaktions-Systems betrachten. Eine formale Anwendung des Algorithmus für das Berechnen der elementaren Flussmoden (Routen) erbringt das Resultat, dass die Zahl solcher Moden manchmal vom Niveau der Beschreibung abhängt, wenn einige Reaktionen reversibel sind und die Produkte der multifunktionellen Enzyme externe Metabolite sind, oder einige multifunktionelle Enzyme zum Teil die gleichen Stoffwechselprodukte umwandeln. Jedoch kann dieses Problem durch eine geeignete Deutung der Definition der elementaren Moden und die korrekte Wahl der unabhängigen

Funktionen der Multifunktionsenzyme gelöst werden. Die Analyse wird durch einige kleinere Beispiele und ein größeres biochemisches Beispiel veranschaulicht, das aus dem Nukleotidmetabolismus stammt und die zwei Arten der Beschreibung für Nukleosiddiphosphokinase und Adenylatekinase vergleicht.

Der Nukleotidmetabolismus spielt eine wichtige Rolle in lebenden Organismen und ist gegenüber allen möglichen Störungen in seiner internen Balance sehr empfindlich. Gefährliche Krankheiten können auftreten, wenn einige Enzyme nicht richtig funktionieren. Mit Hilfe des Konzeptes des elementaren Flussmodus erklären wir das Auftreten und den Schweregrad von Krankheiten, die auf Enzymdefizienzen basieren. Wenn ein Enzym vollständig gehemmt wird, werden einige metabolische Wege blockiert. Wenn jedoch einige alternative Wege noch bestehen, ist die Krankheit weniger gefährlich. Unsere Analyse ist darauf gerichtet, alternative Wege, wesentliche Enzyme und solche Enzyme, die immer zusammenarbeiten zu finden. Der letzte Begriff ist auch als ,,*Enzyme subset* " bekannt und stellt einen intermediären Schritt im Algorithmus zur Berechnung der elementaren Flussmoden dar. Wir diskutieren bereits bekannte und bisher nur hypothetische Mechanismen einiger Krankheiten (proliferative Krankheiten, Immundefizienzen), die auf Störungen des Nukleotidmetabolismus oder seiner Ausbeutung durch Viren und Parasiten beruhen. Die meisten Strategien, die für das Bekämpfen solcher Krankheiten eingesetzt werden, basieren auf der Unterbrechung des Nukleotidmetabolismus an bestimmten Stellen. Diese Strategien können aber auch zur Akkumulation toxischer Stoffe führen und dadurch Nebenwirkungen hervorrufen. Deswegen hilft ein besseres Verständnis dieses Systems, wirkungsvollere Medikamente zu entwickeln, und eine gute strukturelle Analyse kann viele experimentelle Bemühungen ersparen.

Konzepte aus der Theorie der Petri-Netze liefern zusätzliche Werkzeuge für das Modellieren metabolischer Netzwerke. In Kapitel 4 werden die Ähnlichkeiten zwischen einigen Konzepten in der traditionellen biochemischen Modellierung und analogen Konzepten aus der Petri-Netztheorie besprochen. Zum Beispiel entspricht die stochiometrische Matrix eines metabolischen Netzwerkes der Inzidenzmatrix des Petri-Netzes. Die Flussmoden und die Erhaltungs-Relationen haben die T-Invarianten beziehungsweise P-Invarianten als Gegenstücke. Wir decken die biologische Bedeutung einiger weiterer Begriffe aus der Theorie der Petri-Netze auf, nämlich ,,*traps*", ,,*siphons*", ,,*deadlocks* " und ,,*Lebendigkeit*". Wir konzentrieren uns auf der topologischen Analyse anstatt auf die Analyse des dynamischen Verhaltens. Die geeignete Behandlung der externen Stoffwechselprodukte wird ebenfalls besprochen. Zur Illustration werden einige einfache theoretische Beispiele vorgestellt. Außerdem werden einige Petri-Netze präsentiert, die konkreten biochemischen Netzen entsprechen, um unsere Resultate zu belegen. Zum

Beispiel wird die Rolle der Triosephosphatisomerase (TPI) im Metabolismus von *Trypanosoma brucei* ausgewertet, indem *traps* und *siphons* ermittelt werden. Alle behandelten Eigenschaften von Petri-Netzen werden anhand eines Systems illustriert, das aus dem Nukleotidmetabolismus stammt.

Während viele Bemühungen für das Zerlegen metabolischer Systeme, (elementare Flußmoden, extreme Wege) erfolgt sind, sind bisher unseres Wissens keine Versuche in dieser Richtung für Signalübertragungssysteme unternommen worden. Eine spezielle Eigenschaft von Signalnetzwerken in lebenden Zellen ist, dass Aktivierungen, Hemmungen und biochemische Reaktionen normalerweise gleichzeitig anwesend sind. Selbst wenn sie nicht Reaktionen enthalten, machen Mehrfach-Aktivierungen oder Mehrfach-Hemmungen die Netzwerke in hohem Grade verzweigt. Es ist eine schwierige und sehr zeitraubende Aufgabe, alle Faktoren, die einen Einfluss auf ein gegebenes Ziel haben, ohne eine automatische Methode zu ermitteln. Bereits in Kapitel 1 heben wir die Ähnlichkeiten und Unterschiede zwischen den metabolischen und Signal-Netzwerken hervor. In Kapitel 5 errichten wir einen Rahmen und präsentieren einen Algorithmus für die Zerlegung von Signalnetzwerken in kleinere Einheiten, die einfacher zu studieren und zu verstehen sind. Zwei Fälle werden untersucht: ein einfacheres, wenn nur monomolekulare Aktivierungen oder Reaktionen anwesend sind, und ein komplizierterer Fall, wenn die Aktivierungen und die Reaktionen multimolekular sein können. Ihre Beschreibung erfordert unterschiedliche Methoden: klassische Graphen bzw. Petrinetze. Wir besprechen die Probleme, die in unserem Modell wegen des Vorhandenseins von Hemmungen oder von unbekannten Effekten im Netz auftreten. Der vorgeschlagene Algorithmus ermittelt die Faktoren, die zusammenwirken und die Zielsubstanzen, die auf dem gleichen Weg beeinflusst werden. Die Zyklen, die im System auftreten, und mögliche fehlende Reaktionen werden ebenfalls ermittelt . Theoretische Beispiele veranschaulichen unsere Resultate. Anhand der T-Zell-Antigen-Rezeptor-Signalkaskade zeigen wir, wie die Methoden in realen Systemen angewendet werden können.

Schlagwörter:

Enzyme niedriger Spezifität, multifunktionelle Enzyme, Halbreaktionen, Enzymdefizienzen, Nukleotidstoffwechsel, proliferative Krankheiten, Immunodefizienz, Petri-Netze, Elementarmoden,Signalnetzwerk, Signaltransduktion, Computermodellierung.

Summary

The living organisms are too complex when considering them as a whole. The present thesis deals with the topological properties of two important parts of living organisms: the metabolic and the regulatory systems. The topological properties are those features that are determined by the network structure. A classification in metabolic and regulatory systems is often used. A signalling system is a special kind of regulatory system. Between metabolic and signalling networks, there are important differences that impose their treatment in different ways. In metabolic pathway analysis, the elementary flux mode concept is already established as a proper tool for identifying the smallest essential routes in biochemical systems. We examine its features and advantages in some particular cases.

Firstly, many enzymes operate with low specificity (e.g. nucleoside diphosphokinase, uridine kinase, transketolase, transaldolase), so that various substrates and products can be converted. Also the enzymatic mechanisms are diverse, as we have illustrated with reaction schemes for bifunctional enzymes. Therefore, there are two different approaches to describe multifunctional enzymes (We considered only the case when a certain active site hosts several reactions. The case when the studied enzyme has several such active sites can be transformed into that of several enzymes having only one active site. If a disease alters the initial enzyme, also all substituting enzymes are altered.): in terms of overall reactions and in terms of reactions steps (hemi-reactions, half-reactions). For enzymes with two or more functions, it is important to consider only linearly independent functions, because otherwise cyclic elementary modes would occur which do not perform any net transformation. However, the choice of linearly independent functions is not *a priori* unique. In Chapter 2, we give a method for making this choice unique by considering the convex basis of the hemi-reactions system. The set of linearly independent functions provided by our algorithm coincides, in the case of transketolase in pentose phosphate pathway, with the set of linearly independent functions mentioned in literature. A formal application of the algorithm for computing elementary flux modes (pathways) yields the result that the number of such modes sometimes depends on the level of description if some reactions are reversible and the products of the multifunctional enzymes are external metabolites or some multifunctional enzymes partly share the same metabolites. However, this problem can be solved by appropriate interpretation of the definition of elementary modes and the correct choice of independent functions of multifunctional enzymes. The analysis is illustrated by a biochemical example taken from nucleotide metabolism, comparing the two ways of description for nucleoside diphosphokinase and adenylate kinase,

and by several smaller examples.

The nucleotide metabolism plays an important role in living organisms and is very sensitive to any perturbations in its internal balance. Dangerous diseases may occur if some enzymes do not work properly. With the help of elementary flux mode concept, we explain the occurrence and severity of diseases based on enzyme deficiencies. If an enzyme is completely inhibited, some metabolic routes are blocked. If, however, some alternative routes still exist, the disease is less dangerous. In chapter 3, we focus on finding alternative routes, essential enzymes and enzymes operating together. The latter notion is also known as ,,enzyme subset" and represents an intermediary step in calculating the elementary flux modes. The known or hypothesised mechanisms of several disorders, occurred due to the malfunctioning of nucleotide metabolism (proliferative diseases, immunodeficiency diseases) or due to its hijacking by viruses and parasites, are given. Most strategies adopted for curing such diseases are based on nucleotide metabolism interruption. Therefore, a better understanding of this system helps developing more effective drugs and a good structural analysis can spare many experimental efforts.

Petri net concepts provide additional tools for the modelling of metabolic networks. In Chapter 4, the similarities between the counterparts in traditional biochemical modelling and Petri net theory are discussed. For example, the stoichiometry matrix of a metabolic network corresponds to the incidence matrix of the Petri net. The flux modes and conservation relations have the T-invariants, respectively, P-invariants as counterparts. We reveal the biological meaning of some notions specific to the Petri net framework (traps, siphons, deadlocks, liveness). We focus on the topological analysis rather than on the analysis of the dynamic behaviour. The treatment of external metabolites is discussed. Some simple theoretical examples are presented for illustration. Also the Petri nets corresponding to some biochemical networks are built to support our results. For example, the role of triose phosphate isomerase (TPI) in *Trypanosoma brucei* metabolism is evaluated by detecting siphons and traps. All Petri net properties treated in above-mentioned chapter (4) are exemplified on a system extracted from nucleotide metabolism.

While for decomposing metabolic systems, many efforts have been done (elementary flux modes, convex basis, extreme pathways), for signalling maps, as far as we know, no attempt in this direction has been made. A special characteristic of signalling networks is that activations, inhibitions, and biochemical reactions are normally present in parallel. Even if they do not contain reactions, multi-part activations or inhibitions make them highly branched. To detect all factors that have an influence on a given target, without using an automatic method, is a difficult and very time-consuming effort. Already in Chapter 1 (Backgrounds), we highlight the similarities and dif-

ferences between metabolic and signalling networks. In Chapter 5, we build a framework and algorithm for decomposing signalling networks in smaller units, which are easier to study and understand. Two cases are investigated: a simpler one, when only monomolecular activations or reactions are present, and a more complex case, when the activations and reactions can be multimolecular. Their description requires different instruments: classical graphs and Petri nets, respectively. We discuss the problems that occur in our model due to the presence of some inhibitions or unknown effects in the network. The algorithm that we propose detects the factors that are acting together and the targets that are affected on the same route. The cycles that occur in the system are also highlighted. We point out possible missing reactions. Theoretical examples illustrate out findings. Using the T cell antigen-receptor signalling cascade, we show how it can be applied to real systems.

Contents

List of Figures

List of Tables

Chapter 1

Overview

1.1 Background

Approaches in the modelling of metabolic networks

Living organisms have developed continuously and reached a great diversity. Through experimental methods, perseverant scientists have elucidated complex metabolic and regulatory networks. A huge variety of models have been elaborated in theoretical biology to facilitate the study of real systems. As any model, each framework that was defined has some restrictive assumptions. Two important branches can be distinguished up to now in biochemical research – the investigation of metabolic networks and of regulatory networks [HS96, Men97, SKW00].

Several approaches can be distinguished in studying biochemical networks:

- kinetic modelling,

- metabolic control analysis,

- structural approach,

- optimisation.

The *kinetic modelling* is exemplified by studies like those elaborated by [MS88, Gol96]– regarding signalling pathways - and [HHDG85, TWvDW98, WPS$^+$00], investigating glycolysis. [KB73, HR74, Red88, MHTSK90, BHB90, Fel92] explore metabolic and regulatory networks using *metabolic control analysis*. The *structural approach* can be illustrated by the studies done

by [PSVNn$^+$99, SDF99, FW00, JTA$^+$00, SLP00, OBJO&D01]. The evolution of metabolic pathways in the light of *optimisation* was highlighted by [FS86, MH99, HSH91, MCS98, MMHC99].

Thus, biochemical reaction networks are the subject of extensive modelling studies. A number of *kinetic models* (e.g. [MS88, GDB90]) have been proposed to explain the periodic calcium spiking induced by a constant stimulus. [WPS$^+$00] developed a method to establish by which route dynamics propagate through a biological reaction network and applied it on yeast glycolysis, where the concentrations of metabolites oscillate under some conditions. [Gol02] gave an overview of computational approaches that might help the understanding of oscillatory behaviour (calcium oscillations, pulsatile intercellular communication, and circadian rhythms) in biological systems.

Another model, developed and applied by [HHDG85] to calculate time-dependent states of energy transduction in isolated rat liver mitochondria, provided a deeper insight into function and regulatory structure of the systems than obtainable by studying only the stationary states. Their idea was also supported by the fact that in many experiments on isolated mitochondria, non steady-state conditions are met.

[TWvDW98] stressed the importance of the negative feedback acting on their first steps for the so-called turbo design pathways. The lack of this "guard" at the gate of glycolysis leads to the phenomenon of substrate-accelerated death, caused by the impossibility to reach a steady state.

The bases of *metabolic control analysis* (MCA) were independently set by [KB73] and [HR74]. They provided a method to analyse the sensitivity of a metabolic system to perturbations of the environment or of the internal state of this system. Linear enzymatic chains were analysed at steady state. Simple, analytical solutions are found for the metabolite concentrations and the flux through the chain. To describe the enzyme systems, three new concepts (control strength, control matrix, and effector strength) were proposed.

[Red88] emphasised the importance of investigating the structural features of a system from the point of view of metabolic control analysis. The properties of control matrices were highlighted and generalizations of the summation and connectivity theorems were given. [MHTSK90] clarified the definition of transition time of a metabolic system and its relation with other important systemic properties, as flux and concentration control coefficients. They also derived a summation and a connectivity theorem. [BHB90] proposed a "top-down" approach to determine the flux and concentration control coefficients. [Fel92] reviewed the metabolic control analysis indicating the scope of basic theory, its modifications and proposed extensions, experimental applications and areas of disagreement.

[KHWB97] developed a framework for quantitative analysis of a signal

transfer through cellular signal transduction pathways and networks, based on the formalism that was first developed for metabolic networks. Differences between the behavior of signalling cascades and metabolic networks were highlighted.

After early attempts by [KW91], [DlFSWM02] extended MCA to hierarchical control analysis (HCA). This extension was necessary due to the connection existing between metabolic pathways and signal-transduction and gene expression, which has as consequence that enzyme activities are not generally constant, as MCA assumed.

Following a *structural approach*, [FW00] found similarities in the structure of metabolic and sociological networks and highlighted structural properties that could inform us about the network's evolutionary history. They identified the center of the small world of *E. coli* metabolism in glutamate and pyruvate, excluding from the study common coenzyme as ATP, ADP or NAD, which are connected by a very small number of steps to all other metabolites. Taking the possible way of generating such small worlds into account, they deduced that glycolysis and tricarboxylic acid cycle could be the most ancient metabolic pathways. They also emphasized that such a design principle allows the metabolism to react rapidly to perturbations.

[JTA$^+$00] investigated the large-scale organization of metabolic networks. Comparing 43 organisms, they found out that some metabolites (always the same) act as hubs. This leads to a constant networks' diameter for all studied organisms, which leads to an unexpected robustness against random errors.

Optimisation. [MH99] showed the role that the hypothesis of simplicity played in the evolution of pathways, modelling the pathway structure of the pentose phosphate cycle as a mathematical game of combinatorial optimisation. The optimal conversion of pentoses into hexoses must follow the enzymatic mechanisms and requires the least number of steps to the products and the least number of carbons in every intermediate. The Calvin cycle and L-type pentose phosphate cycle provided other example to support the same idea.

[HSH91] outlined the mathematical approaches used to study the evolution and structural design of biochemical reaction systems from the point of view of optimisation principles. They proposed various ways of expressing biological constraints resulting in "cost functions" in mathematical terms, although this is sometimes difficult. Also other sources of difficulties (consideration of systemic properties, nonlinearity of the optimisation problems, etc) are discussed.

A nice example of optimisation, stressing not the pathway structure, but the structure of a macromolecule, was given by [MMHC99]. They proved the fractal structure of glycogen. They not only showed the mathematical func-

tions based on fractal geometry algorithms describing the glycogen synthesis and degradation, but revealed also the biological meaning of this iterative process, establishing the advantages of such a design and pointing out the constrains that ruled the model.

Other studies ([TTK95, BI99, SKW00, HNR02, KKB$^+$02, SPA$^+$02, WSL03] draw attention to the signalling cascades, which are a special type of regulatory systems.

"Divide et impera"– Decomposing biological networks

But even considering only metabolic networks, the explored subject is huge. Therefore, it was necessary to find a proper decomposition in parts that can be treated more easily. Intuitively, only relatively small parts of the metabolism were studied: glycolysis and gluconeogenesis, Krebs cycle, the Calvin Cycle and the pentose phosphate pathway, nucleotide bioynthesis ([Kar77, WAL87, Str95]), etc. Their isolation from the whole cell metabolism was based on specific functions: anaerobic degradation of glucose to pyruvate with production of ATP and glucose synthesis from noncarbohydrate precursors in cytosol, the analogous aerobic processes which take place in mitochondria, photosynthesis, de novo synthesis of nucleotide and their salvage pathways. But the basic unit was not yet found. Writing the flux pathways at steady state as a linear superposition of "fundamental modes" was used by [LB87] to analyse the apparent wasting of free energy (in terms of ATP equivalents) in substrate cycles (called also futile cycles) and "dual pathways"(i.e., parallel routes). This is a first step to a proper decomposition of the metabolic networks.

Pathway analysis has become an important tool in biochemical modelling [SAN98, JTA$^+$00, SLP00, SFD00, RB01, VDL02]. It is also helpful in functional genomics [BDDL$^+$98, DSS$^+$99, Pal00, SP00] and biotechnology [LHC96, SDF99, SFD00, SPM$^+$02, VDL02].

The declared aim of [VDL02] was to direct experimental approaches for metabolic reconstructions, to evaluate possible alternatives of the network, to determine what resources are limiting the cell potential and to suggest how these resources can be directed away from biomass and towards the synthesis of the desired products. [RB01] developed an approach based on both pathway analysis and kinetic modelling to asses the biochemical control of sucrose accumulation and futile cycling in sugar cane. According with their findings, overexpression of the fructose or glucose transporter or the vacuolar import protein, as well as reduction of cytosolic neutral invertase levels, appear to be the most promising targets for genetic manipulation.

A central concept in pathway analysis is that of elementary flux modes

[SH94, SDF99, SFD00]. An elementary mode is a minimal set of enzymes that could operate at steady state. "Minimal" means that if only the enzymes belonging to this set were operating at steady state, complete inhibition of one of them would stop any steady-state flux in the system. If we write a flux distribution as a vector, $\mathbf{V}$, with the components corresponding to the fluxes through the particular enzymes, then this definition implies that $\mathbf{V}$ represents an elementary mode if there is no other steady-state flux distribution, $\mathbf{V}$', having zero components wherever $\mathbf{V}$ does and at least one additional zero component [SHWF02]. The reversible reactions are treated each as a single reaction that can operate both forwards and backwards. No spurious cycles occur because reversible reactions are not decomposed into 2 different reactions each.

Another definition says that an elementary mode cannot be decomposed into two flux modes involving fewer enzymes [SH94, HS96],

$$\mathbf{V} \neq \lambda_1 \mathbf{V}' + \lambda_2 \mathbf{V}", \quad \lambda_1, \lambda_2 > 0$$

It has recently been proved that these two definitions are equivalent [SHWF02]. Any flow pattern can be expressed as a non-negative linear combination of the elementary modes. In this way, we can find, for a given biochemical system, the metabolic routes that lead from a particular starting metabolite to the desired reaction product, and it becomes clear whether a particular enzyme is essential in the biotransformation. Elementary modes have been determined, for example, for sucrose metabolism in sugar cane [RB01] and for carbon metabolism in *Methylobacterium extorquens* [VDL02].

Related concepts in pathway analysis were proposed [LHC96, NnSVPI+97, SB88, SLP00]. [SLP00] proposed the concept of extreme pathways for decomposing the metabolic systems. They also aspired to analyse, interpret and predict the metabolic functions from a pathway-based perspective in addition to the traditional reaction-based perspective. According to this approach, each reversible reaction is split into two irreversible reactions having opposite directions. This leads to occurrence of biochemical irrelevant cycles among the computed elementary pathways.

Another concept used to decompose the metabolic networks is that of the convex basis [NnSVPI+97, PSVNn+99]. This is a subset of the set of elementary modes with the property that no basis vector can be expressed as a superposition of other basis vectors. Elementary modes and the convex basis can be computed by the C program METATOOL [PSVNn+99], which is available from

http://www.bioinf.mdc-berlin.de/projects/metabolic/metatool/.

Once the definition of elementary flux modes was given, the pathways defined intuitively in biochemistry until now (such as glycolysis) are used to exemplify the advantages of this structural analysis. New features can be investigated. The optimal conversion yields and prediction of the effects of the insertion or deletion of enzymes were exemplified on central carbon metabolism [SKB+02]. Among others, properties occurred due to a transformation of irreversible reactions in reversible ones were pointed out by [SHWF02].

Multifunctional enzymes

Many enzymes are known to be multifunctional. This means that different substrates might be bound in one or more than one active sites. We focus on those enzymes having only one active site, but with a low specificity for the substrates. An example is provided by uridine kinase. It can phosphorylate uridine, cytidine and a number of derivatives of these nucleosides (such as 5-fluorouridine) [LA75, LGMA75]. As phosphate donors, various nucleoside triphosphates can be used. Competitive inhibition by alternative substrates binding to the same active site need not be taken into account in computing the elementary modes because these modes refer to potential flux distributions. The magnitudes of reaction rates and the percentage of each elementary mode with which it contributes to the actual flux distribution are not the point of our interest in pathway analysis.

The description of multifunctional enzymes can be made in two different ways. The usual way is by writing overall reactions in terms of substrates and products. In the context of pathway analysis, an alternative was proposed by [NnSVPI+97]. This approach describes the enzymatic mechanism in more detail considering the steps of the formation, interconversion and decay of the enzyme-substrate complex (half-reactions, hemi-reactions). Such a detailed description has also been used in other areas of biochemical modelling [LGMA75].

Apart from the study by [NnSVPI+97], up to now, multifunctional enzymes have been treated in pathway analysis at the level of overall reactions. Of course, the different functions of such an enzyme then have to be treated as distinct reactions when computing the elementary modes (for example, by the program METATOOL). For example, the two functions of transketolase have been labelled as TktI and TktII in [SFD00] (see also Fig. 1.1). For the computation of pathways, this description implies that the two reactions are treated independently. However, if knockout mutations or enzyme deficiencies are studied, care has to be taken that the two reactions are deleted simultaneously from the list of reactions. The choice of these independent
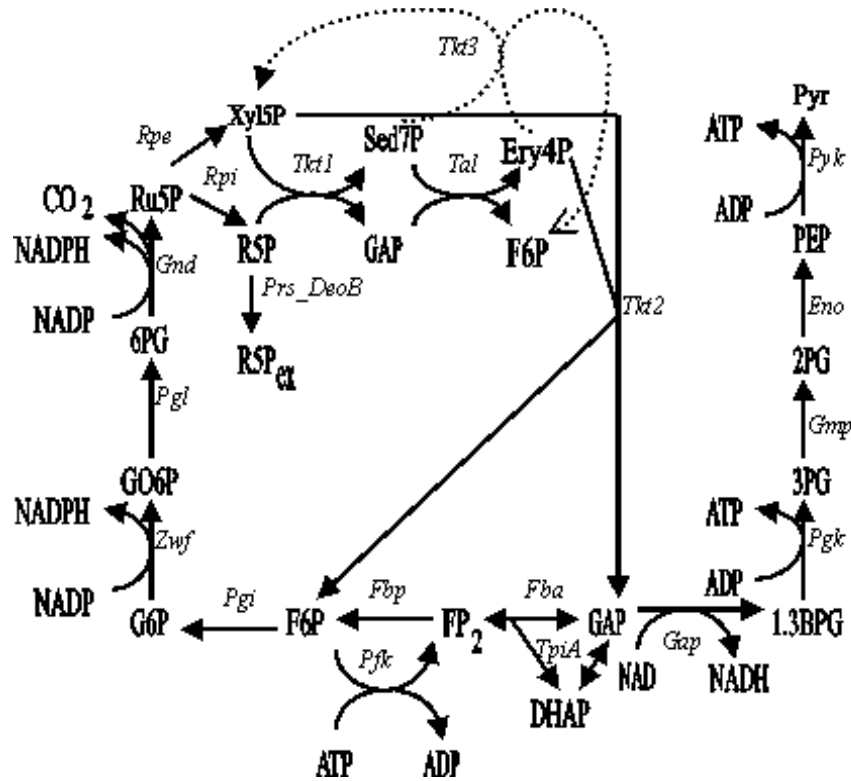
Figure 1.1: Reaction scheme of the glycolysis/pentose phosphate pathway (PPP) system. All the enzymes shown occur in the cytosol of many cells. Tkt1, Tkt2, and Tkt3 stand for the three functions of transketolase acting on sugars except eight-carbon sugars. **Metabolite names:** ADP: adenosine diphosphate; ATP: adenosine triphosphate; 1.3BPG: 1,3-biphosphoglycerate; DHAP: dihydroxy-acetonephosphate; Ery4P: erythrose 4-phosphate. F6P: fructose 6-phosphate; FP2: fructose 1,6-biphosphate; G6P: glucose 6-phosphate; GAP: glyceraldehyde 3-phosphate; GO6P: 6-phosphogluconolactone; NADH/NAD: nicotinamide adenine dinucleotide (reduced and oxidized forms); NADPH/NADP: nicotinamide adenine dinucleotide phosphate (reduced and oxidized forms); PEP: phosphoenolpyruvate; 2PG: 2-phosphoglycerate; 3PG: 3-phosphoglycerate; 6PG: 6-phosphogluconate; Pyr: pyruvate; R5P: ribose 5-phosphate; R5Pex: ribose 5-phosphate moiety incorporated in nucleotides; Ru5P: ribose 5-phosphate; Sed7P: Sedoheptulose 7-phosphate; Xyl5P: xylulose 5-phosphate; **Enzyme names:** Eno: enolase (EC: 4.2.1.11); Fba: fructose 1,6 biphosphate aldolase (EC: 4.1.2.13); Fbp: fructose 1,6 biphosphatase (EC: 3.1.3.11); Gap: glyceraldehyde 3-phosphate dehydrogenase (EC: 1.2.1.12); Gnd: phosphogluconate dehydrogenase (decarboxylating) (EC: 1.1.1.44); Gpm: phosphoglycerate mutase (EC: 5.4.2.1); Pfk: 6-phosphofructokinase (EC: 2.7.1.4); Pgi: phosphoglucoisomerase (EC: 5.3.1.9); Pgk: phosphoglycerate kinase (EC: 2.7.2.3); Pgl: phosphogluconolactonase (EC: 3.1.1.31); Prs_DeoB: 5-phosphoribosyl-1-pyrophosphate synthetase/phosphopentomutase; yk: pyruvate kinase (EC: 2.7.1.40); Rpe: ribose 5-phosphate epimerase (EC: 5.3.1.6); Rpi: ribulose-phosphate 3-epimerase (EC: 5.1.3.1); Tal: transaldolase (EC: 2.2.1.2); Tkt1/2/3: transketolase (EC: 2.2.1.1); TpiA: triosephosphate isomerase (EC: 5.3.1.1); Zwf: glucose 6-phosphate dehydrogenase (EC: 1.1.1.49);

functions is usually not unique. But selecting an arbitrary independent functions set can lead to wrong flux modes that are not at all elementary. The following questions arise: Can the right independent functions be found? Is it helpful to consider a more detailed level of description? Does the number of elementary flux modes and the set of modes depend on the level of description?

Enzyme deficiencies

Nucleotide metabolism contains many enzymes with low substrate specificity. Also here, it is known that many diseases occur as a consequence of enzyme deficiencies, that is, the insufficient expression of an enzyme. The diseases mainly have two reasons: The missing or incompletely operating enzyme can imply the interruption of a synthesis pathway and, thus, to the lack of product, or it can lead to the accumulation of an intermediate that is toxic when present in high concentration or inhibits another essential enzyme. Once the cause of the disease is identified, it becomes possible to supply the enzyme that is insufficient (for a good functioning of the organism) in the form of a drug. Care has to be taken because the enzymes are usually macromolecules. If they would be orally administered, they are cut to peptides by the enzymes in the digestive tract. Therefore a way to avoid this route has to be found, for example, injection into the blood [HM00]. Another cause of the diseases based on enzyme problems is a mutations series in the DNA sequence that is responsible for the synthesis of the enzyme in question. Due to these mutations, the enzyme may lose its binding or catalytic properties. Independent of the cause that produced the defective functioning of the enzyme, the effect could be disastrous for the patients.

The importance of nucleotide metabolism is well known. Nucleotide triphosphates are the substrates for RNA and DNA synthesis. This offers some points for control of cancer [SBS91, DSS$^+$99, FG99, MKK01, CLW02], viral infections [GJCM95] and parasites [FB02]. The hydrolysis of nucleotide triphosphates fuels many metabolic reactions. Several nucleotides (GTP, cAMP, etc.) serve as regulatory molecules. Nucleotides are also components of many coenzymes, such as coenzyme A, NAD, and FAD.

There are two possibilities to synthesize nucleotides: One is by the *de novo* pathways, and the second, by the salvage pathways, which recycle fragments of the nucleic acids. There are some cells, such as erythrocytes and many parasitic species such as pathogenic protozoa and *Mycoplasma,* where only salvage routes operate, because *de novo* systems do not exist at all. This makes the cells vulnerable to pharmaceutical intervention. Also in neurons, there is a poor *de novo* synthesis.

An important enzyme in nucleotide metabolism is HPRT (hypo-xanthine-guanine phosphoribosyl transferase). Its deficiency leads to gout. A complete loss of HPRT (HPRT, EC 2.4.2.8) is the cause for the Lesch-Nyhan syndrome. In this case, the patients present spasticity with pyramidal tract signs, compulsive self-mutilation, choreoathetosis and developmental retardation [JF00].

The occurrence of diseases based on enzyme deficiencies will be explained with the help of the concept of elementary flux modes in Chapter 3. We will tackle questions such as: why are some enzymes defects lethal while others cause only less dangerous diseases? Are there several enzymes whose deficiencies could lead to the same disease? Can enzyme deficiency diseases be cured?! In which measure pathway analysis can help in the case of proliferative and immunodeficiency diseases? Can the diseases caused by viruses and parasites be compared with enzyme deficiency diseases? Is this comparison of any help?

The effect of enzyme deficiencies have been studied by mathematical modelling earlier, notably for the energy and redox metabolism in erythrocytes [SJH89, SH95a]. In those studies, kinetic modelling rather than pathway analysis has been used.

Can Petri Net Theory help in modelling metabolic networks?

Specific features of biochemical systems are that most reactions are catalysed by enzymes and that many reactions utilize more than one substrate (reactant) and/or generate more than one product. Reaction systems that only involve isomerizations (that is, reactions with one substrate and one product) can be depicted as graphs and their properties can be studied from the point of view of graph theory. However, real biochemical networks cannot be represented as graphs due to bi- or multimolecular reactions. These cases would require that arcs linking three or more nodes exist. One, quite complicated, approach to coping with this situation is to consider the groups of substances on each side of reaction arrows as nodes (so-called complexes) [HJ72, Cla80].

A simpler solution is offered by Petri net theory [Rei85, Sta90]. Two kinds of nodes are considered: places and transitions. The nodes and arcs between them represent the static structure, while some more elements/components such as tokens indicating time-dependent weights of places are used to describe the dynamics. Beside place/transition nets, also condition/event Petri nets have been proposed in the literature. Which approach is proper to model biochemical systems?

Petri nets can be employed for the graphical description of processes.

They allow us to understand more intuitively the temporal evolution of systems by considering flows of tokens through the nets. They offer also an appropriate formalism for the analysis of biochemical networks, as has been pointed out earlier by several authors [Hof94, LGMA75, HKS00, KSYA00, HKV01, GKV01, Pal00, OBJO&D01], while the above-mentioned modelling approaches [HS96, LB87, TWvDW98, FW00, JTA$^+$00, SHWF02, SKW$^+$02] are independent of Petri net theory. Are the models based on Petri nets totally new, or can one find the same concepts in other clothes ? Do the concepts of Petri nets theory, which have not yet a correspondence in traditional metabolic network analysis, have however a meaning from the point of view of living processes?

Comparison between metabolic and signalling networks

During biological evolution, living cells have developed mechanisms for adaptation to environmental changes. In recording these variations, membrane receptors play an important role. They are sensitive to the presence of some chemical compounds (Ca^{2+}, Mg^{2+} ions, hormones, regulatory proteins, etc.) and also to variation of temperature, pressure, vibration and so on. The information about what is happening outside of the cell is sent inside as a signal and propagated in the cell via complex signalling cascades and networks. The modelling of such networks is a growing field [TTK95, BI99, HNR02, KKB$^+$02, SPA$^+$02, WSL03]. An external impulse triggers a chain of events inside. This means that different cellular substances are set in motion and interact with each other. They form complexes, activate other metabolites or complexes, catalyse important reactions, or inhibit others so as to prepare a proper response to the change in the cell's environment. These phenomena are depicted in very suggestive signalling maps. But these maps usually get very complicated when the number of the agents involved increases. Sometimes, the substances are in both activated and inactivated forms and also the supposed intermediary steps (e.g. transport across a membrane) for achieving the contact with the next substance are shown. We shall simplify the description as much as possible, considering only one representation for each substrate in a cell compartment. When a substrate X has to cross a membrane between two compartments A and B, we shall represent it by a reaction between two different substrates (X_A and X_B), as is usually done in biochemical modelling. The interactions carrying the signals are represented by arrows.

The substances are classified into initial factors, intermediaries and targets. Initial factors are those substances that initiate the signal. They could be regulatory proteins, ions, hormones or the like. Temperature or mechan-

ical vibrations can be also considered as factors. It is only a matter of notation. The endpoints of the system under consideration are called targets. They can also reside in the nucleus (genes, for example). The intermediaries are the substances in between. The distinction between initial factors, targets and intermediaries depends on the definition of the model. When the system is extended, almost each factor or target may become an intermediary, while on reducing the system some intermediaries could become targets or initial factors.

Signalling mechanisms were developed during evolution as an adaptation machinery to environmental changes. Due to the huge number of proteins involved in these cascades, this machinery is very complex. Approximately half of the 25 largest protein families encoded by the human genome deal primarily with information processing [BTS02]. Due to their large polarity and dimensions, most signal molecules are not able to diffuse through the cell membrane. Therefore, the signals are transmitted to other molecules inside the cell, such as cAMP, Ca^{2+}, inositol-trisphosphate (IP_3), and diacylglycerol (DAG). These bind to proteins that may interact with DNA and alter the gene-expression pattern. Gene expression is a complex phenomenon, in which many regulatory proteins are involved. Once the signal penetrates the cell membrane, is amplified and a response is generated, it has also to be terminated. Without termination, cells would lose their responsiveness to new signals, or, even worse, an uncontrolled growth will lead to cancer.

There are important differences between signalling networks and metabolic networks. While the latter are characterized by a conversion of substrates into products and, hence, by mass flow, signals are usually transferred without a mass flow from the starting point to the target. For the activation and inhibition effects in signalling networks, no mass balance condition need to be fulfilled. For example, some MAPKK molecule can catalyse the phosphorylation of several MAPK molecules without being consumed at all. Accordingly, signalling networks usually enable signal amplification, which is, in a sense, incompatible with mass conservation. At a molecular level, there are conversions going on, for example, phosphorylations. However, in enzyme cascades, the phosphate moiety is not channelled along the entire signalling route (cf. [HNR02]. (In the phosphotransferase system, though, the phosphate moiety is channelled, cf. [PL85]) Therefore, we cannot usually consider a signal as a continuous reaction flux. Moreover, the temporal aspect plays a greater role in signalling networks than in metabolic networks, which often subsist in steady states. As for signalling networks, two cases can be distinguished. There can be a switch from one steady state to another or the systems are inherently dynamic. In the former case, methods of metabolic control analysis can be used to quantify the amplification of signals [KHWB97, SFD00]. Ex-

amples are provided by long-acting hormones such as the thyroid hormone. In the dynamic case, which is much more difficult to describe in quantitative terms, time-dependent signals, for example, short pulses are propagated. Examples are provided by hormones with short periods of action such as adrenaline. Interestingly, also mixed forms exist. For example, a permanently elevated level of a hormone such as phenylephrine (switch to another steady state) may cause the onset of oscillations in the intracellular Ca^{2+} level in hepatocytes (cf. [Gol96]).

The connections between substances have several types. First, the interaction can be a binding process. For example, in the B cell antigen-receptor signalling cascade, the factors Grb2, Sos, and Vav and the enzyme PLCγ bind to form a complex [Cam99]. Such processes can be described in the same way as biochemical reactions. Another type of interaction is the covalent modification of a substance, the regulatory or catalytic properties of which are thus changed. The best known modification is phosphorylation, for example, by MAP kinase. Moreover, methylation, acetylation, adenylylation, uridylylation and other covalent modifications occur. Depending on whether the modified form of the substance is more active or less active than the unmodified form, one speaks of activation or inhibition, respectively. Unfortunately, it often happens in signalling networks that we know that a metabolite can affect another metabolite, but whether it has a positive or a negative effect is still unknown. In this case, we will simply call it "effect".

Activations, inhibitions, effects or biochemical reactions are gathered together in large, complex networks. Although activations and inhibitions are, at a molecular level, also reactions, for example, phosphorylations mediated by kinases, they are often represented in a simpler way, using Boolean algebra and handling changes in the "logical state" of proteins: inactive or active. But it is not always possible to represent a reaction as an effect, for example, a dimerisation. Two molecules of the same metabolite may have to bind to each other to activate another substance. Therefore, signalling maps may consist only of reactions, if a biochemical representation at a detailed level is adopted, or they can mix effects and reactions.

Decomposition of signalling networks into functional units

To better understand a large, intricate signalling map, it is helpful to decompose it into simpler parts. Analogously, metabolic networks are often decomposed into simple functional units. Widely used definitions of such units are that of elementary flux modes [SH94, SFD00] and the similar concept of extreme pathways [SLP00]. These approaches are not immediately applicable for the decomposition of signalling networks because of the above-

mentioned differences between signalling and metabolic networks. Nevertheless, we here suggest that an analogous and specifically adapted approach is valuable for modelling signalling networks. The basic idea is to detect all non-decomposable routes through the network. We can say that a signalling route is only balanced if at each intermediary, there is one incoming effect and one outgoing effect, quite similarly as it is in elementary modes in biochemical networks.

From the point of view of signal transduction, the functional units that we look for are striking. We would like to know which initial factor(s) can bring about changes in (a) given target(s). Hence, we will search all the routes ending at a given target. Due to the usually high degree of branching, it is possible that not only one factor is acting on our target, but a set of factors. It can occur that a subset of initial factors are working together in all the routes obtained. Another result could be that several targets are, always or only through some routes, hit in parallel. One could be interested in discovering those targets that are influenced by changing a given initial factor. Consequently, we will also identify the routes starting with a given factor. We are also interested in detecting parallel routes that start at the same initial factor and end at the same target. Such redundant routes are important for network robustness. In metabolic networks, robustness due to network redundancy has been a subject of intense studies [JTA$^+$00, SKB$^+$02].

Our aim is to deduce as much information as possible from the structure of the network under consideration. Kinetic parameters are not to be included, because they are not or only incompletely known for most signalling networks.

1.2 Motivation, aims and organization

This thesis aims to give answers to the questions that have been raised in the previous sections. For a given a system, already from its structure one can derive important properties. Importantly, all these topological features that we study are invariants. The term ,,topology" is here used in the sense of ,,structure of a reaction network", that is, in which way substances are connected with each other by reactions. Considering also some parameters, as the kinetics in the dynamical modelling or the initial markings in the Petri nets representation, the model comes alive. We focus on the topological properties because they are much better known than kinetic parameters and because they are still not completely exploited.

Regarding metabolic systems, for which a decomposition tool already exists, it is important to show that this instrument is invariant to the level of description. Calculating the elementary flux modes for big models could

need many resources in terms of time and memory. Considering a more detailed description of a given system enlarges, in facts, the input data, so the requirements of time and memory. Therefore, it is valuable to ignore the enzymatic mechanism without impairing the result. A problem still remains. Since the choice of these independent functions is usually not unique and selecting an arbitrary independent functions set can lead to wrong flux modes, that are not at all elementary, how can this be avoided? In the next chapter (2), an answer will be given and it will be supported with an algorithm. Here, the half-reactions will play an essential role. Since for finding the independent functions fewer resources are needed, although it is necessary to take into account the enzymatic mechanism, it is not so expensive any more. The example, taken from the nucleotide metabolism will illustrate the analysis.

The concept of elementary flux modes can help to clarify the cause of dangerous diseases. Essential enzymes can be recognized and, if they work deficiently, they or their activators can be supplied with drugs, in an appropriated way. Possible alternative routes for a given function can be discovered and then activated by drugs, or artificial alternative routes can be built, using over-expressing methods of biotechnology, so that diseases based on enzymes deficiency can be alleviated. The severity degree of some enzymes defects can be predicted and its occurrence - avoided. The third chapter (3) will illustrate these findings on nucleotide metabolism. As this part of the metabolism is normally responsible for proliferative and immunodeficiency diseases, and also exploited by viruses and parasites as well, we highlight the similarities and the differences between these three cases and how they are reflected into cure strategies. Although the diseases related with this metabolism part are already almost known, matching results obtained by two different methods (experimental and theoretical) represents a step forward. They should be applied also on unexamined systems.

An alternative way of modelling, which can partially overlap with the traditional one, but which can also bring new aspects is brought under a magnifying-glass in the forth chapter (4). Applied only to metabolic systems, our analysis is done on place/transition Petri nets. Later on, when the signalling networks will enter the scene, we shall deal with condition/event coloured Petri nets [Jen98]. In the forth chapter (4), we shall show the correspondence between concepts in both, Petri nets theory and traditional metabolic network analysis. Some examples will help the reader see the similarities and to exploit them. We shall focus on the topological analysis of biochemical networks by using Petri nets rather than on the analysis of the dynamic behaviour. In particular, we shall deal with various invariants and other features in these nets such as dedlockfreeness and liveness and re-

veal their biochemical meaning. Moreover, we shall discuss the appropriate treatment of source and sink metabolites.

Coloured Petri nets approach can be also useful to design signalling networks. The similarities and differences between metabolic and signalling networks will be pointed out and a proper framework to decompose them into functional units will be built. Depending on the complexity degree of the networks, two models will be presented. One corresponds to signalling networks containing only monomolecular reactions and effects. This model is based on classical graphs and allows presence of both activations and inhibitions as effects. Thus, applying the algorithm that we proposes, all route starting in a given initial factor and leading to each possible reachable target will be detected and their total effect will can be evaluated. The second model handles with more complex networks consisting of multipart activations and multimolecular reactions. Using a backtracking-based algorithm, the simple routes between special sets of initial factors and targets, routes that are not any more sequences, will be detected. An example from B cell antigen receptor signalling pathway illustrates the idea.

Chapter 2

Treatment of multifunctional enzymes

2.1 Introduction

Due to their capacity to catalyse biochemical reactions, the enzymes are important actors in metabolic world. Most of them are very specific, being able to recognize and accept the mediating of only those substrates involved in one single reaction. But other many enzymes have active sites characterized by a low specificity. Thus, they are able to facilitate several reactions.

Such multifunctional enzymes are known to be of huge importance, being often responsible for vital processes in organism, their deficiencies frequently leading to severe diseases. Let us view several multifunctional enzymes acting in human sugar metabolism: Transketolase is an essential enzyme in the pentose phosphate pathway (PPP) [MH99]. It is involved in the synthesis of ribose-5-phosphate as a precursor of nucleotides and erythrose-4-phosphate as a precursor of aromatic amino acids. Transketolase has been analysed in detail in cancer research. It is supposed that the growth of tumour cells with their highly active nucleotide synthesis could be reduced by inhibition of this enzyme [CCV$^+$00]. Tkt is extremely abundant in human white blood cells. A loss of its activity occurs in several diseases (e.g. chronic lymphatic leukaemia); Tkt is also present in high levels in mammalian cornea. NADPH produced by the PPP has a role in the removal of light-generated radicals [UUY$^+$00]. Another multifunctional enzyme is transaldolase. It is involved in the synthesis of sugars, sugar analogues and related compounds. Like Tkt, it has a key role in the conversion of R5P into glycolytic intermediates. A Tal deficiency could produce liver cirrhosis and persistent hepatosplenomegaly; there is also a hypothesis saying that HIV-induced apoptosis could be reg-

ulated by Tal [VHR+01]. Many diseases are associated with an increased presence of aldolase, such as muscular dystrophy, acute hepatitis and other liver diseases, myocardial infarction and prostate cancer [RKL+87].

Tkt, Tal and Ald are reversible enzymes, being able to act both forward and backwards, depending on the substrates availability. Most of the time, the reactions directionality complicates the model. Therefore, let us first consider the case where all multifunctional enzymes involve only irreversible steps.

2.2 The case where all multifunctional enzymes involve only irreversible steps

For uridine kinase, the reaction with the highest activity is

Urk1: ATP + uridine = ADP + UMP.

In the following, we list all reactions phosphorylating uridine that have more than about 50% of the activity of this reaction as well as all reactions phosphorylating cytidine that have more than about 50% of the activity of the reaction

Urk2: ATP + cytidine = ADP + CMP,

(which, in turn, has 35.6 % of the activity of Urk1) (see database BRENDA, http://www.brenda.uni-koeln.de/).

Urk3: dATP + uridine = dADP + UMP,
Urk4: dATP + cytidine = dADP + CMP,
Urk5: dUTP + uridine = dUDP + UMP,
Urk6: dUTP + cytidine = dUDP + CMP,
Urk7: dCTP + uridine = dCDP + UMP,
Urk8: dCTP + cytidine = dCDP + CMP.
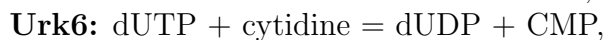
In every cell containing a given multifunctional enzyme, all its different functions are performed simultaneously, but to avoid redundancy, we will consider only independent functions. The other ones can be expressed as linear combinations of these. For example, Urk4 = -Urk1 + Urk2 + Urk3. The number of linearly independent reactions is given by the rank of the stoichiometry matrix [Alb94, HS96]. This can be calculated either by standard tools from linear algebra or by determining the number of metabolites minus

the number of independent conservation relations. We here apply the second method. There are twelve metabolites involved in uridine kinase: ATP, dATP, ADP, dADP, UMP, CMP, dUTP, dUDP, dCTP, dCDP, uridine, cytidine, and seven conservation relations, notably for the moieties ADP, dADP, dUDP, dCTP, uridine, cytidine, and phosphate. Thus, the number of independent reactions is five (12 - 7 = 5). As in any vector space, the choice of basis vectors is not unique. Here, one could choose, for example, the following five reactions: Urk1, Urk2, Urk3, Urk5, Urk7. Two of these are shown in Fig. 2.1.
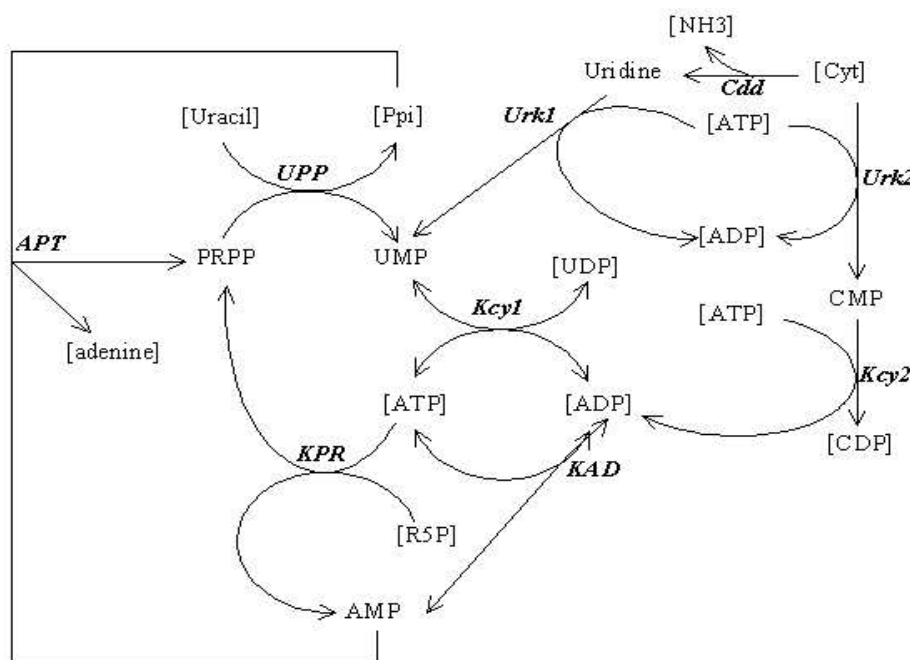
Figure 2.1: Uridine kinase system. Enzymes: **APT**, adenine phosphoribosyltransferase (EC 2.4.2.7); **Cdd**, cytidine deaminase (EC 3.5.4.5); **KAD**, adenylate kinase (EC 2.7.4.3); **Kcy1**, cytidylate kinase (EC 2.7.4.14); **KPR**, ribose-phosphate pyrophosphokinase (EC 2.7.6.1); **UPP**, uracil phosphoribosyltransferase (EC 2.4.2.9); **Urk**, uridine kinase (EC 2.7.1.48). Metabolite names: ADP, adenosine diphosphate; AMP, adenosine monophosphate; ATP, adenosine triphosphate; CDP, cytidine diphosphate; CMP, cytidine monophosphate; Cyt, cytidine; NH3, ammonia; Ppi, diphosphate; PRPP, phosphoribosylpyrophosphate; R5P, ribose-5-phosphate; UDP, uridine diphosphate; UMP, uridine monophosphate. The arrow with double arrow-head corresponds to the production of two molecules of metabolite.

As mentioned in Chapter 1, multifunctional enzymes can be described in two different ways. The formulation of a reaction in terms of elementary steps (half-reactions) depends on the reaction mechanism. For example, a bifunctional reaction, A + B → C + D and F + B → G + D, can proceed according to the different mechanisms shown in Fig 2.1. There are opposing views in the literature about the mechanism of uridine kinase [LA75, LGMA75, Ore68, PCT85, see also database BRENDA]. Let us first assume that it operates according to an ordered sequential mechanism [LGMA75, PCT85]. Clearly the phosphate donor binds first. For symmetry reasons, this implies that (d)NDP is released last (with N standing for A, U, or C). Then, the half-reactions read:

ATP + Urk = Urk-ATP (i.e. the enzyme-substrate complex)
Urk-ATP + uridine = Urk-ADP + UMP
Urk-ATP + cytidine = Urk-ADP + CMP
Urk-ADP = Urk + ADP
dNTP + Urk = Urk-dNTP
Urk-dNTP + uridine = Urk-dNDP + UMP
Urk-dNTP + cytidine = Urk-dNDP + CMP
Urk-dNDP= Urk + dNDP.
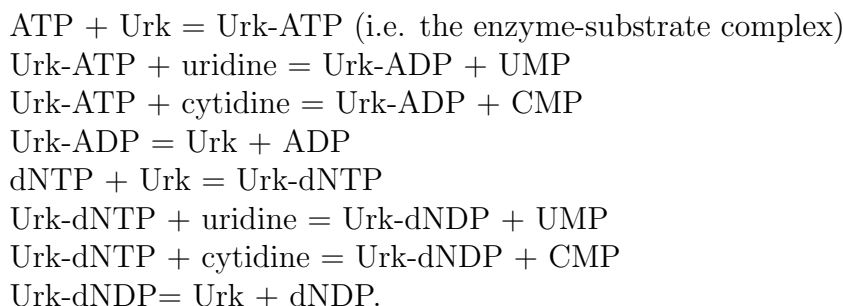
More rigorously, one should write the formation and decay of ternary complexes such as Urk-ATP-uridine. However, for establishing the input to METATOOL, it is not necessary to write all elementary steps. Consecutive steps without branching in between can be lumped into one step because any elementary mode containing one of these steps will also include the other [PSVNn+99]. For example, in the ordered ping-pong mechanism indicated in Fig. 2.2(a), we can lump the first two steps (E + B → EB, EB → D + EP) into E + B → D + EP and the last two steps A + EP → EC, EC → C + E into A + EP → C + E, respectively F + EP → EG, EG → E + G into F + EP → E + G.

If the enzymatic mechanism of uridine kinase is ordered ping-pong [Ore68], the half-reactions read:

ATP + Urk = Urk-P + ADP   dCTP + Urk = Urk-P + dCDP
dATP + Urk = Urk-P + dADP   Urk-P + uridine = UMP + Urk
dUTP + Urk = Urk-P + dUDP   Urk-P + cytidine = CMP + Urk.
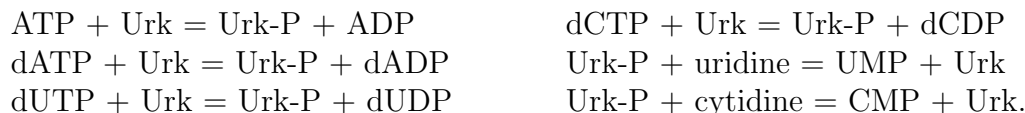
We consider a system made up of uridine kinase and several other reactions of nucleotide metabolism (Fig. 2.1). Here, the number of elementary

modes does not depend on the level of description nor on the enzyme mechanism. In each case, there are four elementary flux modes, which in terms of overall reactions read: {Kcy2, Urk2}; {Kcy1, Cdd, Urk1}; {UPP, Kcy1, KAD, KPR}; {(2 UPP), (2 Kcy1), KPR, APT}. All of them are irreversible. Moreover, the convex basis here coincides with the set of elementary modes.

2.3 The case where some multifunctional enzymes involve reversible steps

A linear dependence among the various reactions of a multifunctional enzyme only occurs if these reactions share partly the same substances. For example, an enzyme catalysing the reversible reactions A ↔ B and A ↔ C will also catalyse the reaction B ↔ C. In contrast, there is no linear dependence for an enzyme catalysing the reactions A ↔ B, C ↔ D. The term "bifunctional enzyme" should be used for an enzyme with two linearly independent functions.

A prominent example of a reversible multifunctional enzyme is transketolase (EC 2.2.1.1). It is a homodimer, but each monomer can be studied separately without restricting the analysis. Each of its monomers can catalyse the reactions [Str95] **Tkt1:** ribose-5P + xylulose-5P = sedoheptulose-7P + glyceraldehyde-P, **Tkt2:** erythrose-4P + xylulose-5P = fructose-6P + glyceraldehyde-P (where P stands for phosphate). Interestingly, a linear combination can be written: **Tkt3:** erythrose-4P + sedoheptulose-7P = ribose-5P + fructose-6P (-Tkt1, Tkt2), which is also a bimolecular reaction. Usually, in the literature, only Tkt1 and Tkt2 are mentioned [SHL+83, Str95, YO80]. However, one can likewise indicate, for example, Tkt1 and Tkt3 as independent functions of transketolase. It is important to realize that the choice of independent functions is, up to now, somehow arbitrary. Accordingly, one cannot assign net fluxes to all the different functions of a multifunctional enzyme on the basis of the knowledge of the production or consumption rates of the metabolites converted by this enzyme.

Moreover, it was hypothesized that transketolase can also act on octulose-8-phosphate (O8P) [NnSVPI+97, WAL87]. This would imply the following additional reactions:

Tkt4: O8P + GAP = F6P + X5P .
Tkt5: 2 F6P = E4P + O8P .
Tkt6: S7P + F6P = R5P + O8P .
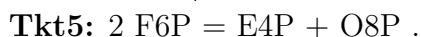
However, the latter two reactions are not independent of the others; they can
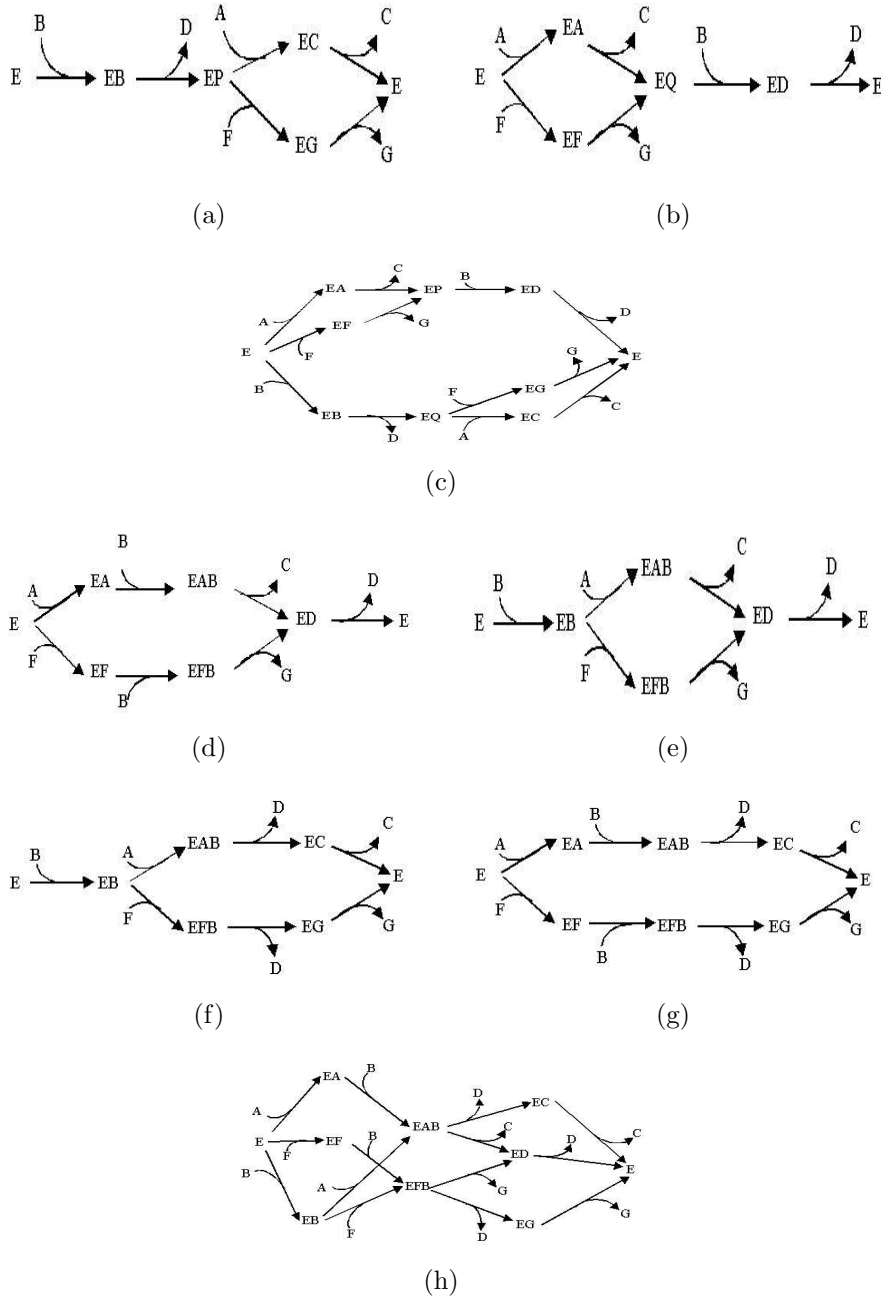
Figure 2.2: Different reaction mechanisms for the bifunctional reaction A + B → C + D and F + B → G + D. (a) Ordered ping-pong with low specificity in the lower part. (b) Ordered ping-pong with low specificity in the upper part. (c) Random ping-pong. (d)-(g) Ordered sequential with low specificity in various steps. (h) Random sequential.

be written as Tkt5 = (-Tkt2, -Tkt4); Tkt6 = (-Tkt1, -Tkt4). We consider reactions Tkt1, Tkt2 and Tkt4 as a basis for the others.

As the transketolase reaction follows the ordered ping-pong mechanism [FGS$^+$01], it can be described as:

R1 : C5 + TK = C3 + TKC2 .   **R3** : C7 + TK = C5 + TKC2 .
R2 : C6 + TK = C4 + TKC2 .   **R4** : C8 + TK = C6 + TKC2 .

We used the following notation referring to the number of carbon atoms:

GAP =: C3,   E4P =: C4,
X5P or R5P =: C5,   F6P =: C6,
S7P =:C7,   O8P =: C8

because the main feature of the transketolase functions is to carry two atoms of carbon (the so-called glycolaldehyde residue) from one reactant to another. TKC2 stands for the enzyme-substrate complex consisting of transketolase and a C2-unit.

We start with studying a reaction system which is very similar to the system investigated in [NnSVPI$^+$97]. It involves sugars with three to eight carbons. Not only transketolase but also transaldolase (with two independent functions) and aldolase (with three independent functions) are here considered multifunctional. This system was proposed to be relevant for erythrocyte and liver metabolism [WAL87].

Using the abridged notation introduced above, the system can be written as follows:

Functions of transketolase: as given above**.**

Functions of transaldolase: Functions of aldolase:
Tal1: C4 + C6 = C3 + C7   **Ald1:** C6 = 2C3
Tal2: C4 + C8 = C7 + C5   **Ald2:** C7 = C3 + C4
Tal3: C8 + C3 = C6 + C5   **Ald3:** C8 = C3 + C5
(Tal3=Tal2 - Tal1)
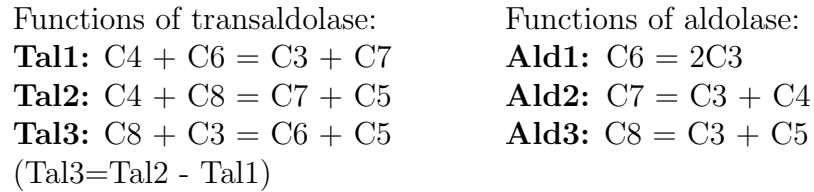
The system under study also has some more reactions:

OPPP : C6 + 2 NADP = C5 + CO2 + 2 NADPH
(lumped reaction sequence of oxidative PPP)
RCon : C5 = C5ex (ribose consumption)
Glucim : C6ex = C6 (glucose import)
PyrCon : C3 = C3ex (pyruvate consumption).
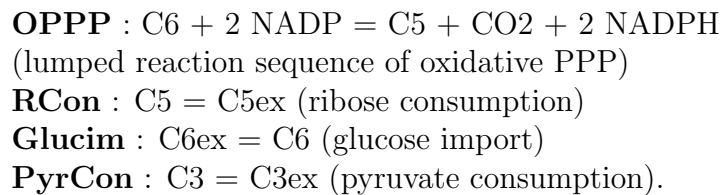
Name	Decomposition
C3	C3
C4	C4
C5	C3-C2
C6	C4-C2
C7	C3-C2-C2
C8	C4-C2-C2

Table 2.1: The composition of the metabolites participating in the transketolase reactions.

Name	Decomposition
C3	C3
C4	C4
C5	C5
C6	C3-C3'
C7	C4-C3'
C8	C5-C3'

Table 2.2: The composition of the metabolites participating in the transaldolase and aldolase reactions.

When considering only linearly independent reactions (e.g. only Tal1 and Tal2, but not Tal3), the numbers of elementary flux modes and basis vectors are, using METATOOL, computed to be 37 and 6. If also linearly dependent reactions were included, we would obtain elementary modes representing reaction cycles. For example, when including not only Tkt1 and Tkt2 but also **Tkt4:** C4 + C7 = C5 + C6, we obtain 86 elementary modes. One of these is {-Tkt1, Tkt2, Tkt4} which does not perform any net transformation of substances. Moreover, there are several modes which differ from other modes just by a replacement of {-Tkt1, Tkt2} by {Tkt4}. To avoid irrelevant cyclic modes, it is appropriate to consider only linearly independent functions of multifunctional enzymes.

As mentioned in the in the previous section, to determine a set of linearly independent reactions, one can use standard methods from linear algebra or, alternatively, one can calculate them in a less abstract way. This will be illustrated here for the system under study. The reaction subsystem made up of the six functions of transketolase involves six metabolites and three conservation relations, namely for C2, C3 and C4. Note that C5, C6, C7 and C8 are composed of C3 and C2, C4 and C2, C3 and two C2 moieties,

and C4 and two C2 moieties, respectively (Table 2.1). Note that always the moieties of which the metabolite is composed and which are not decomposed in the system, are conserved. Thus, the number of linearly independent functions is 6 - 3=3. For transaldolase, four conservation relations exist, notably for the C3, C4 and C5 moieties which can exist both freely or bound and the persistent C3 moiety (denoted C3' in Table 2.2) which can only exist in the bound state. Thus, the number of linearly independent functions of transaldolase is 6 - 4=2. Table 2.2 shows the composition of the metabolites participating in the transaldolase and aldolase reactions.

Transketolase, transaldolase and aldolase operate according to the ordered ping-pong mechanism (database BRENDA,

<div align="center">http://www.brenda.uni-koeln.de/).</div>

Thus, the reaction system reads, in terms of hemi-reactions,

R1 : C5 + TK = C3 + TKC2 .   **R5** : C6 + TA = C3 + TAC3 .
R2 : C6 + TK = C4 + TKC2 .   **R6** : C7 + TA = C4 + TAC3 .
R3 : C7 + TK = C5 + TKC2 .   **R7** : C8 + TA = C5 + TAC3 .
R4 : C8 + TK = C6 + TKC2 .   **R8** : C6 = 2 C3 .

R9 : C7 = C3 + C4 .
R10 : C8 = C3 + C5 .
OPPP : C6 + 2 NADP = C5 + CO2 + 2 NADPH .
RCon : C5 = C5ex .
Glucim : C6ex = C6 .
PyrCon : C3 = C3ex .
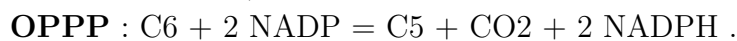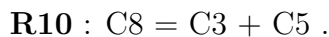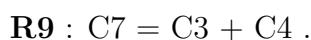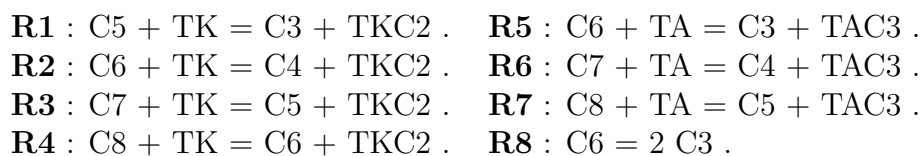
Now, we have more reaction steps than in the overall description. The number of elementary flux modes seams to change to be 65, while the number of basis vectors remains the same (6). Thus, the question occurs whether all these 65 flux modes are really elementary. So, do the elementary flux modes depend on the description of the system? To answer these questions, we first consider a simple example.

2.4 A simple hypothetical example

We will now consider the simple hypothetical system shown in Fig. 2.3(a). Reaction 2 is depicted in terms of hemi-reactions. Half-reaction **b** is reversible, while **a** and **c** are irreversible. Applying the algorithm for computing elementary modes [SFD00, SHWF02] formally to the hemi-reactions system, the following modes are obtained:
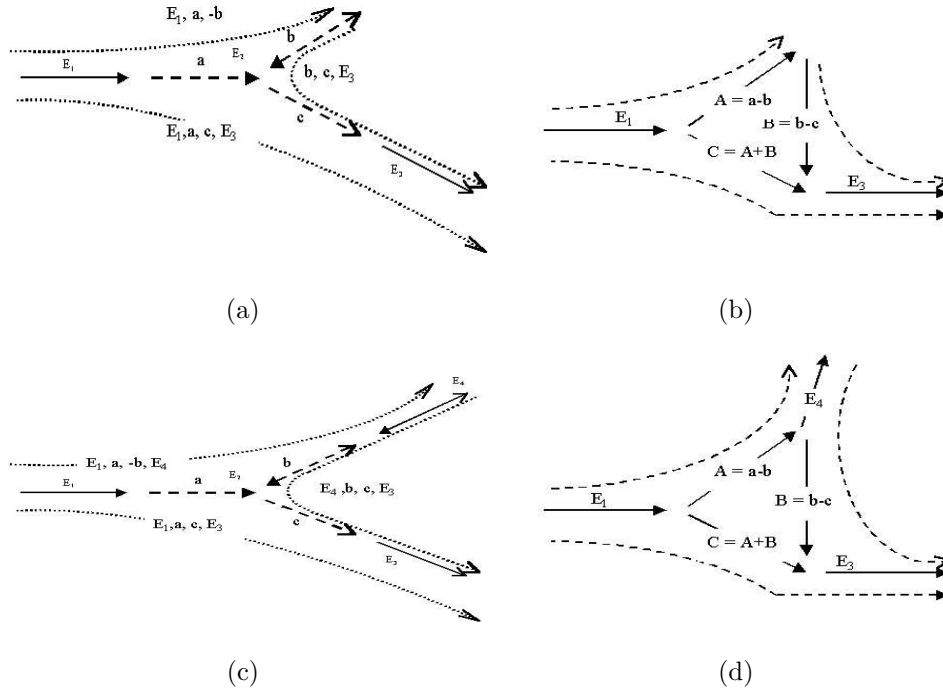
Figure 2.3: Simple reaction schemes illustrating elementary modes and convex basis on different levels of description. **(a)** The multifunctional enzyme is represented by half-reactions (labelled by **a**, **b**, and **c**). Arrows with single (double) arrow-heads refer to irreversible (reversible) reactions. The forward orientation of reversible reactions is symbolized by full arrow-heads. Elementary flux modes are depicted by dashed arrows. **(b)** The same system as in (A) in terms of overall reactions. **(c)** Half-reactions system with an additional enzyme, E_4. **(d)** The same system as in (C) in terms of overall reactions.

$$\{\mathbf{E}_1, \mathbf{a}, \text{-}\mathbf{b}\}, \{\mathbf{b}, \mathbf{c}, \mathbf{E}_3\}, \{\mathbf{E}_1, \mathbf{a}, \mathbf{c}, \mathbf{E}_3\}.$$

The last mode fulfils the conditions mentioned in the two definitions given in the Chapter 1 provided that the components of **V** correspond to reaction steps. However, when steps **a** and **c** are operative, also step **b** can be operative because it belongs to the same enzyme, $\mathbf{E}_2$. Therefore, one can write the third mode as the sum of the first two and, accordingly, it is not elementary.

The corresponding overall-reactions of enzyme E_2 could be, for example:

$$\mathbf{A} = \mathbf{a} \text{ - } \mathbf{b}, \mathbf{B} = \mathbf{b} + \mathbf{c} .$$

The third overall reaction, $\mathbf{C} = \mathbf{a} + \mathbf{c}$, is linearly dependent on **A** and **B**. Therefore, it should not be included on computation of pathways because, otherwise, the meaningless elementary mode $\{\mathbf{A}, \mathbf{B}, \text{-}\mathbf{C}\}$ would occur.

The translation of the above-mentioned modes from the hemi-reactions system in terms of overall-reactions is (Fig. 2.3(b)):

$$\{\mathbf{E}_1,\ \mathbf{A}\},\ \{\mathbf{B},\ \mathbf{E}_3\},\ \{\mathbf{E}_1,\ \mathbf{A}\ +\ \mathbf{B},\ \mathbf{E}_3\}.$$

Now it can clearly be seen that the last one is, according to the definition of elementary flux modes, not elementary, because it is the sum of two modes involving fewer reactions. In the light of the verbal definition of elementary modes given in the Introduction, one may wish to write the modes in terms of enzymes rather than reactions: $\{\mathbf{E}_1,\mathbf{E}_2\}$, $\{\mathbf{E}_2,\mathbf{E}_3\}$, $\{\mathbf{E}_1,\ \mathbf{E}_2,\mathbf{E}_3\}$. However, the last of these is still non-elementary because it involves the former two as subsets.

We will now consider the convex basis for this example system. In the half-reactions description, the basis vectors are $\{\mathbf{E}_1,\ \mathbf{a},\ \text{-}\mathbf{b}\}$ and $\{\mathbf{b},\ \mathbf{c},\ \mathbf{E}_3\}$. The mode $\{\mathbf{E}_1,\ \mathbf{a},\ \mathbf{c},\ \mathbf{E}_3\}$ is not a basis vector because it is the sum of these two pathways. For the given directionality of reactions, the basis vectors are unique. However, if, for example, reactions $\mathbf{a}$ and $\mathbf{E}_1$ were reversible, uniqueness would be lost. We could then choose $\{\mathbf{E}_1,\ \mathbf{a},\ \text{-}\mathbf{b}\}$, $\{\text{-}\mathbf{E}_1,\text{-}\mathbf{a},\ \mathbf{b}\}$ and $\{\mathbf{b},\ \mathbf{c},\ \mathbf{E}_3\}$, or we could choose $\{\mathbf{E}_1,\ \mathbf{a},\ \text{-}\mathbf{b}\}$, $\{\text{-}\mathbf{E}_1,\ \text{-}\mathbf{a},\ \mathbf{b}\}$ and $\{\mathbf{E}_1,\ \mathbf{a},\ \mathbf{c},\ \mathbf{E}_3\}$. In the overall description, if reaction $\mathbf{a}$ is irreversible, the convex basis reads $\{\mathbf{E}_1,\ \mathbf{A}\}$, $\{\mathbf{B},\ \mathbf{E}_3\}$. If reaction $\mathbf{a}$ is reversible, we can take $\{\mathbf{E}_1,\ \mathbf{A}\}$, $\{\text{-}\mathbf{E}_1,\ \text{-}\mathbf{A}\}$, $\{\mathbf{B},\ \mathbf{E}_3\}$ or $\{\mathbf{E}_1,\ \mathbf{A}\}$, $\{\text{-}\mathbf{E}_1,\ \text{-}\mathbf{A}\}$, $\{\mathbf{C},\ \mathbf{E}_3\}$. Thus, in both cases, the number of basis vectors is independent of the level of description. Moreover, for the overall reactions system, the convex basis can be obtained by translating the convex basis of the hemi-reactions system.

As mentioned above, the choice of linearly independent functions of a multifunctional enzyme is not *a priori* unique. In the example under study, if $\mathbf{a}$ were reversible, we could take functions $\mathbf{A}$ and $\mathbf{C}$ instead of $\mathbf{A}$ and $\mathbf{B}$. Formal application of the algorithm then yields the "elementary modes" $\{\mathbf{E}_1,\ \mathbf{A}\}$, $\{\text{-}\mathbf{A},\ \mathbf{C},\ \mathbf{E}_3\}$, $\{\mathbf{E}_1,\ \mathbf{C},\ \mathbf{E}_3\}$. This result is not, however, correct because the third mode is not really elementary. It appears that this choice of linearly independent functions is not suitable. Instead, one should take the functions $\mathbf{A}$ and $\mathbf{B}$. Importantly, the choice of these is derived from the convex basis, $\{\mathbf{E}_1,\ \mathbf{a},\ \text{-}\mathbf{b}\}$ and $\{\mathbf{b},\ \mathbf{c},\ \mathbf{E}_3\}$ (with step $\mathbf{a}$ being irreversible). The convex basis and, thus, the choice of linearly independent functions are imposed on the multifunctional enzyme by the other enzymes in the system.

The above reasoning leads us to establishing a method for identifying the appropriate linearly independent functions.

1. Determine the convex basis in the half-reactions system.

2. See which combinations of half-reactions of the multifunctional enzyme occur and translate these into overall reactions.

3. Compute the elementary modes in the overall reactions system.

In this way, the elementary modes are uniquely defined and do not depend on the level of description. We recall that, on the half-reactions level, the definition of elementary modes should be applied in terms of enzymes rather than of half-reactions.

We now apply the above method to a network describing glycolysis and the pentose phosphate pathway (see [SFD00] for the reaction equations) in which transketolase is described by half-reactions. This gives a convex basis that involves six basis vectors, three of which do not contain transketolase at all, while the others contain 2R1 + R2 + R3. This can be simply translated into Tkt1 + Tkt2. A translation involving Tkt3 would be more complicated, for example, 2 Tkt2 + Tkt3. Thus, the two functions usually given in the literature are in accordance with the above method.

It is interesting to consider the system depicted in Fig. 2.3(c), which extends the system shown in Fig. 2.3(a) in that the substance converted by the reversible hemi-reaction **b** is also converted by an additional enzyme, E_4. For this system the elementary modes in terms of half-reactions read $\{\mathbf{E}_1,$ **a**, -**b**, $\mathbf{E}_4\}$, $\{\mathbf{E}_4,$ **b**, **c**, $\mathbf{E}_3\}$, $\{\mathbf{E}_1,$ **a**, **c**, $\mathbf{E}_3\}$. The last one reads, in terms of overall reactions, $\{\mathbf{E}_1,$ **A** + **B**, $\mathbf{E}_3\}$ (Fig. 2.3(d)). This mode is now elementary because it does not involve enzyme $\mathbf{E}_4$. In this situation, it is thus unnecessary to apply the above method.

2.5   Interconversion of nucleoside triphosphates

To illustrate the main point of this paper, we consider a biochemical system involving the enzyme nucleoside diphosphokinase (EC 2.7.4.6). This enzyme reversibly interconverts various nucleoside triphosphates as well as deoxy-nucleoside triphosphates and the corresponding diphosphates. For simplicity's sake, we here consider only the conversion of ATP, GTP, CTP and UTP. The general reaction equation reads

$N_1TP + N_2DP = N_1DP + N_2TP$ (*)
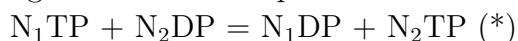
with N_1 and N_2 denoting some nucleosides. As we here consider four different nucleosides, there are 16 different reaction equations of the type (*), of which only six are relevant for symmetry reasons and because N_1 and N_2 should be different.
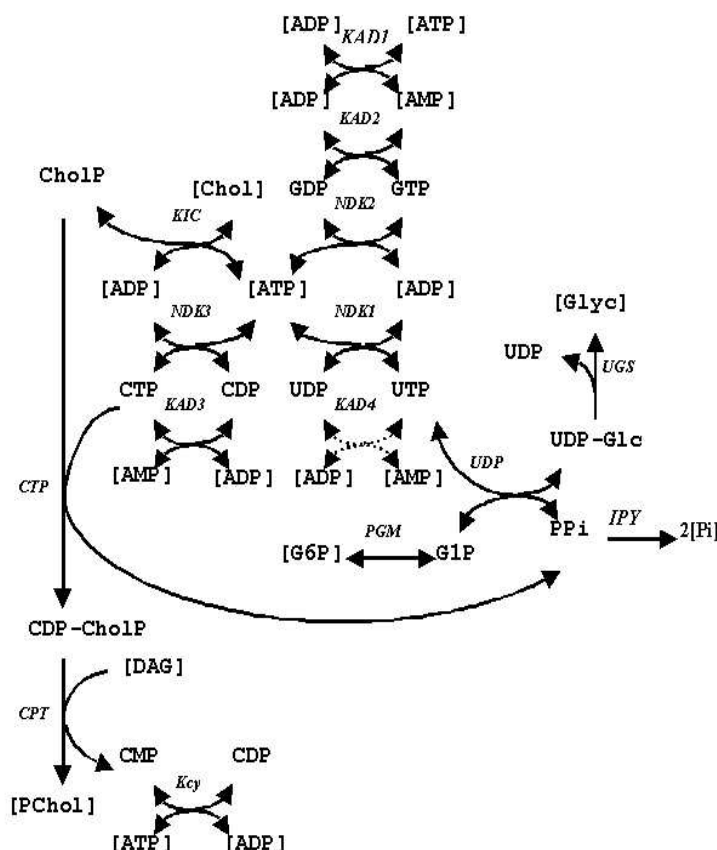
Figure 2.4: Interconversion of nucleoside triphosphates. Enzymes: **NDK**, nucleoside-diphosphate kinase (EC 2.7.4.6); **UDP**, UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9); **PGM**, phosphoglucomutase (EC 5.4.2.2); **KIC**, choline kinase (EC 2.7.1.32); **CTP**, choline phosphate cytidylyltransferase (EC 2.7.7.15); **CPT**, diacylglycerol cholinephosphotransferase (EC 2.7.8.2); **IPY**, inorganic pyrophosphatase (EC 3.6.1.1); **UGS**, glycogen synthase (EC 2.4.1.11); **CYG**, guanylate cyclase (EC 4.6.1.2); **Kcy**, cytidylate kinase (EC 2.7.4.14); **KAD**, adenylate kinase (EC 2.7.4.3). Metabolite names other than those in Fig. 1.1: CDP-Chol, CDP choline; Chol, choline; CholP, cholinephosphate; CTP, cytidine triphosphate; DAG, 1,2-diacylglycerol; G1P, glucose 1-phosphate; G6P, glucose 6-phosphate; GDP, guanosine diphosphate; Glyc, glycogen; GMP, guanosine monophosphate; GTP, guanosine triphosphate; PChol, phosphatidylcholine; UDP-Glc, UDP-glucose; UTP, uridine triphosphate.

Nucleoside diphosphokinase operates according to an ordered ping-pong mechanism [CCV72, see also database BRENDA]. Therefore, in the formalism of half-reactions, the enzyme reaction can be written as

$$\textbf{R1 :} \ ATP + NDK = ADP + NDKP$$
$$\textbf{R2 :} \ GTP + NDK = GDP + NDKP$$
$$\textbf{R3 :} \ UTP + NDK = UDP + NDKP$$
$$\textbf{R4 :} \ CTP + NDK = CDP + NDKP \ .$$
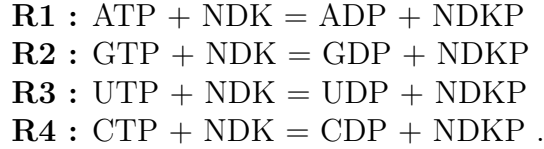
Now we consider the reaction system shown in Fig. 2.4, which extends a network analysed earlier [SPM+02]. For simplicity's sake, we do not indicate transport reactions between different organelles. The network includes, in addition to nucleoside diphosphokinase, several other enzymes of nucleotide interconversion. One of these, adenylate kinase (KAD), is multifunctional as well,

$$\textbf{KAD1} : ATP + AMP = ADP + ADP$$
$$\textbf{KAD2} : GTP + AMP = GDP + ADP$$
$$\textbf{KAD3} : CTP + AMP = CDP + ADP$$
$$\textbf{KAD4} : UTP + AMP = UDP + ADP \ .$$

Its hemi-reactions read:

$$\textbf{R5 :} \ ATP + KAD = ADP + KADP$$
$$\textbf{R6 :} \ GTP + KAD = GDP + KADP$$
$$\textbf{R7 :} \ UTP + KAD = UDP + KADP$$
$$\textbf{R8 :} \ CTP + KAD = CDP + KADP$$
$$\textbf{R9 :} \ ADP + KAD = AMP + KADP \ .$$
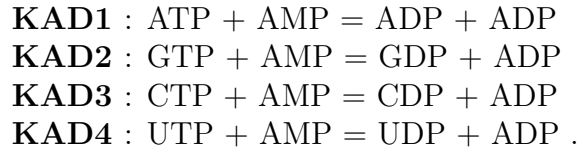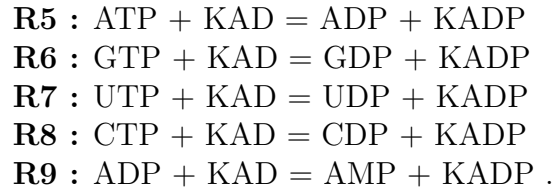
The remaining enzyme reactions are:

$$\textbf{UDP :} \ UTP + G1P = UDP\text{-}Glc + Ppi$$
$$\textbf{CTP :} \ CTP + CholP = CDP\text{-}CholP + Ppi$$
$$\textbf{CPT} : CDP\text{-}CholP + DAG = CMP + Pchol$$
$$\textbf{KIC} : CholP + ADP = Chol + ATP$$
$$\textbf{Kcy} : CDP + ADP = CMP + ATP$$
$$\textbf{PGM :} \ G6P = G1P$$
$$\textbf{IPY :} \ PPi = 2 \ Pi$$
$$\textbf{UGS :} \ UDP\text{-}Glc = Glyc + UDP \ .$$
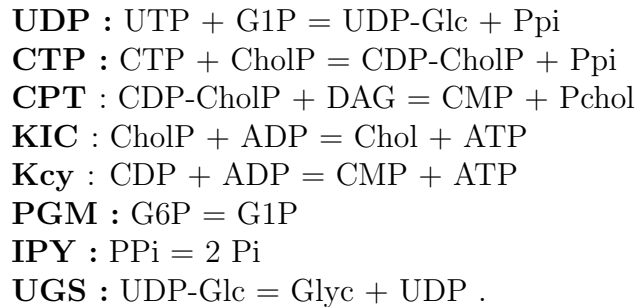
For such a large system, the comparison of elementary modes at the two levels of description is not straightforward. Therefore, we implemented an algorithm which translates the elementary flux modes of the overall-reactions

system into elementary flux modes in terms of hemi-reactions and compares this set with the set of elementary modes computed for the hemi-reactions, identifying the difference. For writing this program we used Lex (Lexical Analyser Generator), with host language C. Our program is available upon request.

According to the method given in Section 3, we first determined the convex basis for the half-reactions system. This involves six vectors. However, three of these are irrelevant because they represent cyclic flows and perform no net transformation (e.g. **R1**, -**R2**, -**R5**, **R6**). The vectors of the convex basis contain the following combinations of hemi-reactions belonging to the same enzyme: (**R1**, -**R2**), (**R1**, -**R3**), (**R1**, -**R4**), (-**R5**, **R6**), (**R5**, −**R7**), (**R5**, -**R8**), (**R5**, -**R9**). In this way, we define the linearly independent functions of KAD and NDK:

$$\textbf{NDK1} = \textbf{R1} - \textbf{R3}\text{: ATP + UDP = ADP + UTP}$$
$$\textbf{NDK2} = \textbf{R1} - \textbf{R2}\text{: ATP + GDP = ADP + GTP}$$
$$\textbf{NDK3} = \textbf{R1} - \textbf{R4}\text{: ATP + CDP = ADP + CTP}$$
$$\textbf{KAD1} = \textbf{R5} - \textbf{R9}\text{: ATP + AMP = 2 ADP}$$
$$\textbf{KAD5} = \textbf{R5} -\textbf{R8}\text{: ATP + CDP = ADP + CTP}$$
$$\textbf{KAD6} = \textbf{R5} -\textbf{R7}\text{: ATP + UDP = ADP + UTP}$$
$$\textbf{KAD7} = -\textbf{R5} + \textbf{R6}\text{: ADP + GTP = ATP + GDP .}$$
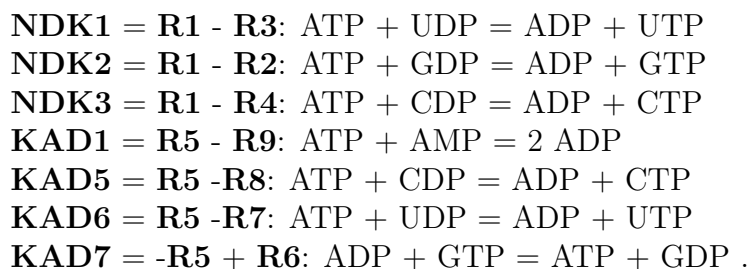
Out of the six reactions of NDK mentioned above, only three are linearly independent because we have eight different metabolites and five conservation relations (for the ADP, GDP, CDP, UDP, and phosphate moieties). Importantly, we have obtained exactly three independent overall reactions from the convex basis. The reaction **NDK4** :GTP + CDP = GDP + CTP, for example, is the sum of the reactions NDK3 and (-NDK2). Note that NDK1, NDK2, and NDK3 are all reversible. As KAD involves one additional metabolite (AMP) and implies the same number of conserved moieties (with AMP instead of ADP being a conserved moiety), the number of linearly independent functions is four. This is in agreement with the functions obtained above.

Table 2.3: List of elementary modes for the reaction scheme of nucleotide metabolism shown in Fig. 2.4.

No	Elementary flux modes in overall reaction system in terms of half-reactions	Elementary flux modes in terms of overall reaction	Directionality
1	-R5 R9	-KAD1	reversible
2	-R1 R2 -R6 R9	NDK2 KAD2	reversible[*]
3	-R1 R3 -R7 R9	NDK1 KAD4	reversible[*]
4	R1 -R3 UDP PGM IPY UGS	NDK1 UDP PGM IPY UGS	irreversible
5	-R7 R9 UDP PGM IPY UGS	-KAD4 UDP PGM IPY UGS	irreversible
6	R1 -R4 R8 -R9	NDK3 KAD3	reversible[*]
7	R1 -R4 -Kcy -KIC CTP IPY CPT	NDK3 -Kcy -KIC CTP IPY CPT	irreversible
8	-R8 R9 -Kcy -KIC CTP IPY CPT	-KAD3 -Kcy -KIC CTP IPY CPT	irreversible

[*] These modes are cyclic and perform no net transformation. They occur because several of the functions of NDK and KAD are identical. They also not elementar, all of them containing the first mode. Therefore, they should be deleted.

Using these functions, computation of elementary modes by the program METATOOL gives eight modes (Table 2.3). However, we have to cancel three of these (modes 2, 3 and 6) because they do not perform any net transformation. Cyclic modes that cannot operate due to thermodynamic reasons are not eliminated by METATOOL. To decide whether these modes can operate, one needs to know whether they are driven by a free-energy difference between external metabolites, which may have been omitted in the reaction equations for the sake of simplicity. It now becomes clear that in the presence of multifunctional enzymes, elementary modes have to be checked carefully whether they realize a net transformation of external substances. Otherwise, non-elementary modes would be obtained. For example, modes 2, 3 and 6 in Table 2.3 include mode no. 1 as a subset. Also in the convex basis, irrelevant cyclic modes have to be cancelled. The number of relevant basis vectors is three in the overall reactions system and, thus, equal to the number obtained above for the hemi-reactions system.

This example shows on a larger scale than the hypothetical example discussed in Section 2.3 how to deal with reversible multifunctional enzymes.

Here, the reason for a special treatment of such enzymes is, however, different. There are sets of substrates of two multifunctional enzymes which overlap. KAD and NDK both use ATP, ADP, UTP, UDP, GTP, GDP, CTP and CDP. Therefore, if we computed the elementary modes for the hemi-reactions system formally (that is, without considering the enzymes as basic units), we would obtain, for example, the mode {R2, -R3, R5, -R6, UDP, PGM, Ppase, UGS}. However, the set of hemi-reactions {R2, -R3, R5, -R6} performs the same overall transformation as the hemi-reactions R1 and -R3, so that this mode is equivalent to mode 4 in Table 2.3.

Chapter 3

Studying enzyme deficiencies by metabolic pathway analysis

3.1 Example from nucleotide metabolism

The system that we studied in the last section of Chapter 2 represents only a small part of the more complex nucleotide metabolism, which contains even more multifunctional enzymes: adenine phosphoribosyltransferase (APRT, EC 2.4.2.7), hypoxanthine-guanine phosphoribosyltransferase (HPRT, EC 2.4.2.8), purine-nucleoside phosphorylase (PNPase, EC 2.4.2.1), 5'-nucleotidase (AMPase, EC 3.1.3.5), cyclic nucleotide phosphodiesterase (cNPDe, 3.1.4.17), apyrase (ADPase, EC 3.6.1.5), xanthine dehydrogenase (Xd, EC 1.1.1.204), xanthine oxidase (XOR, EC 1.1.3.22), etc (see Appendix A). The nucleotide metabolism has been the subject of intense mathematical modelling [JP89, HSG⁺93, SGHS97].

As already said, each function of a multifunctional enzyme has to be initially treated as a different enzyme. But after the elementary flux modes are computed, one has to take into account that if one function of an enzyme does not work properly, the other functions will suffer from the same problem. Therefore, all the elementary flux modes containing functions of an improperly working enzyme, will be affected or even cancelled. Moreover, having the certitude that treating the multifunctional enzymes at the overall reaction level does not lead to misinterpretation, we shall see how pathway analysis can be used in the context of enzyme defects studies.

Let us first consider an overview of the system of interest. The nucleotide metabolism serves to produce nucleosides (purine derivatives: adenosine, guanosine and pyrimidine derivatives: cytidine, uridine and thymidine) and their phosphates (e.g. AMP or CTP), which are called nucleotides, in the

"de novo" pathways. The nucleotides are required to synthesise DNA and RNA and are also used in the synthesis of NAD^+, FAD^+, Coenzyme A, cAMP, cGMP, CDP lipids, ß-alanine, and ß-aminoisobutyric acid [Str95]. The surplus is converted into urea and eliminated. Parts of the bases (A, G, C, U and T) that are not utilised are recycled by the salvage pathways (Fig. 3.1). Usually, taking into account the end-products, one decomposes the nucleotide metabolism into purine and pyrimidine pathways (http://www.genome.ad.jp/kegg/).
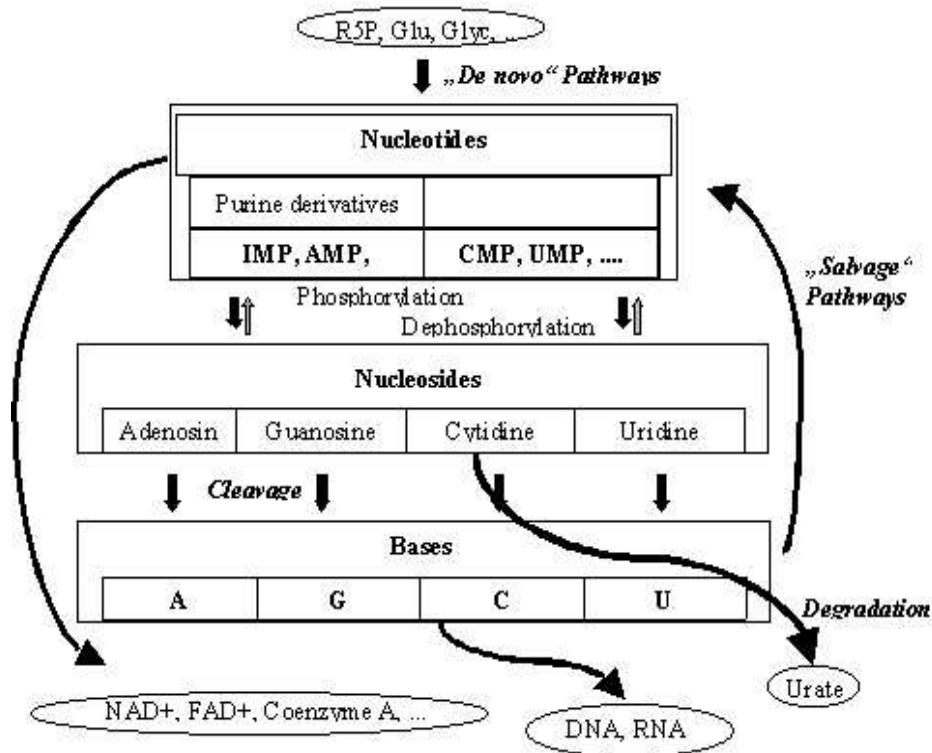
Figure 3.1: Schematic representation of nucleotide metabolism. R5P stands for ribose-5-phosphate, Glu – for L-glutamate, Glyc – for glycine. The other abbreviations have the usual meaning known in biochemistry.

Considering the whole purine metabolism in humans, as it is represented in KEGG

(http://www.genome.ad.jp/dbget-bin/get_pathway?org_name=hsa&mapno=00230),

we have found around 2650 elementary flux modes (depending on which metabolites are considered external). But we think that studying smaller systems presents the advantage of being able to check the solution both

technically and intuitively. Finding on a theoretical way facts that were proven also experimentally, we can develop a method to make predictions for biochemical systems that were not so profoundly studied experimentally. Therefore, we have restricted the initial system to the one given in Fig 3.2.
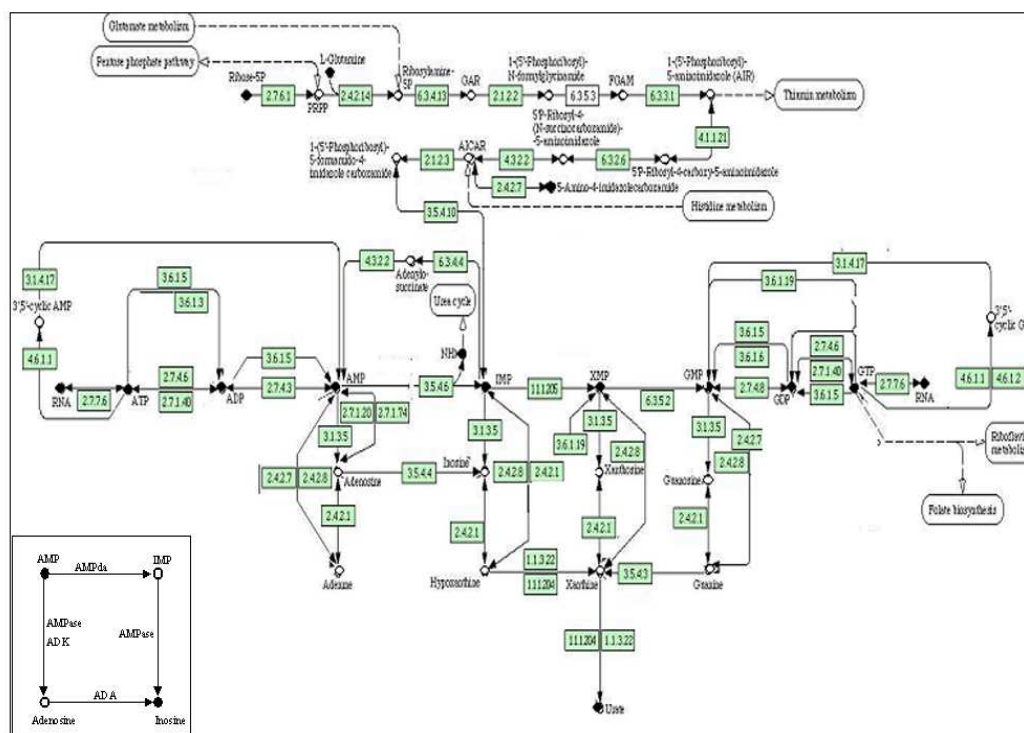
Figure 3.2: Central purine metabolism extracted from KEGG. Empty cycles: internal metabolites, full cycles: external metabolites; grey rectangles: E.C. numbers of the enzymes present in humans. Names and detailed reactions are given in Appendix A. Inset: Schematic representation of the two pararell routes of AMP degradation.

The external metabolites are depicted in full circles. They are: NAD, urate*, NADH2, H_2O_2, NADPH, NH_3*, Pi, PPi, ADP*, ATP*, fumarate, glycine, aspartate, glutamine*, glutamate, pyruvate, PEP, NADP, R5P*, GTP*, GDP*, XMP*, IMP*, AMP*, GMP*. Only those marked with a star are depicted in the scheme extracted from KEGG. In the schemes of that database and in many schemes describing big biochemical systems only one reactant and one product are depicted for each reaction. Other coreactants, considered to be unimportant because they exist in sufficient amount, and the coproducts, which are not further transformed, are ignored. Another reason for overlooking these substances is to simplify the scheme.

There are two criteria for choosing the above-mentioned metabolites as external [SPM$^+$02]. Some of them are boundary metabolites – there are only reactions consuming them (R5P, L-glutamine, etc) or only reactions producing them (urate, NH$_3$). Others are branch metabolites (ATP, GTP, IMP, etc), being used in many reactions, as reactants and also as products. The importance of such branch metabolites will be pointed out in the next section.

It must be said that we studied a modified system than that presented in KEGG. There, many reactions are depicted as reversible, while in the literature [JP89] we find that they are irreversible. Therefore, in our model, APRT, HGPRT and KPY proceed in only one direction.

3.2   Pathway analysis of a purine metabolism model

For this system (see Appendix B), including the enzyme FGAMs (EC 6.3.5.3), converting 1-5'-phosphoribosyl N-formylglycinamide (FGAR) into 2-formami-do-N1-5-phospho-D-ribosyl acetamidine (FGAM), we compute 97 elementary flux modes given in Appendix C, together with their significance. Only one elementary mode (containing FGAMs) stands for a "de novo" pathway of purines, because IMP is set as external. This means that each enzyme which takes part in the elementary flux mode considered above is essential for the "de novo" pathway to function. The conversion between nucleotides is contained in the 82 elementary flux modes, which represent the salvage pathway. 22 of these elementary flux modes lead also to urate. 14 other elementary modes play a role only in nucleotides degradation to urate, producing no additional nucleotides.

In this context, we detected an error in KEGG. In that database, FGAMs is indicated not to function in humans. However, this would imply that there is no "de novo" pathway for purines in humans. This contradicts findings of [BBK$^+$76]. In general, the elementary flux mode concept can be applied to find some gaps and to fill them based on, at least, some suppositions. They then need to be proven later on experimentally.

The "enzyme- or reaction-subsets" for the system under study read as follows: {PNPase3, APRT}, {PNPase5, AMPase6}, {PNPase7, -AMPase4}, {CNPDe, GUc}, {CNPDe2, ADc}, {AIRs2, ASS}, and {IMPc, AIRc, AIRs, SACAIRs, GARt, AICARt, amidoPRT, PRPPs, AIRc2, GARs, FAMs}. As PNPase3, 5 and 6 are three functions of the same enzyme, we can conclude that PNPase functions with APRT and on the other hand with AMPase.

Branch-point metabolites	Enzymes in which the branch-point metabolite is involved
Adenosine	PNPase, AMPase, ADK, ADA
Xanthine	PNPase, Xd, Xd2, XOR, GDA, HPRT2
Hypoxanthine	PNPase, Xd2, HPRT2
Inosine	PNPase2, AMPase2, ADA
Guanine	PNPase7, GDA, HPRT
PRPP	AmoidoPRT, PRPPs, APRT, HPRT2,3,4
R1P	PNPase2, 3, 5, 7

Table 3.1: Branch-point metabolites and the enzymes that use them.

Therefore, when the enzymes (rather than reactions) are considered as the basic units, {PNPase}, {APRT}, and {AMPase} are separate enzymes-subsets. CNPDe and CNPDe2 are two functions of CNPDe. Thus, the two reactions-subsets containing them and only one additional enzyme each will be dissolved into individual subsets {CNPDe}, {Guc} and {ADc}. Again, because AIRs and AIRs2 are two functions of the same enzyme, from the last two subsets given at the beginning of this paragraph, AIRs has to be extracted into individual enzymes-subset. {ASS} and {IMPc, AIRc, SACAIRs, GARt, AICARt, amidoPRT, PRPPs, AIRc2, GARs, FAMs} form new enzymes-subsets. Therefore, the genes responsible for synthesis of the enzymes containing in the latter subset should be expressed simultaneously. The enzymes not belonging to it appear to be synthesised by independently regulated genes. Of course, enlarging the system new constraints could occur, splitting also this subset. Importantly, a defect affects the whole enzyme-subset simultaneously.

Besides most external metabolites, there are also several internal metabolites involved in more than one in- and one out-going reaction. Interventions on the concentrations of such branch metabolites could affect several fluxes in the system. Adenosine participates in 4 reversible reactions, catalysed by PNPase, AMPase, ADK and ADA. Xanthine is involved in one reversible reaction, catalysed by PNPase and 5 irreversible, catalysed by Xd, Xd2, XOR, GDA, and HPRT2. Interestingly, Xd and Xd2 are two functions of the same enzyme, contributing once to the xanthine synthesis from hypoxanthine and once to xanthine degradation to urate. Hypoxanthine, inosine, guanine, PRPP and R1P are also branch-point metabolites and the reactions in which they take part are given in Table 3.1.

In the treatment against severe diseases as hyperuricemia and gout, the
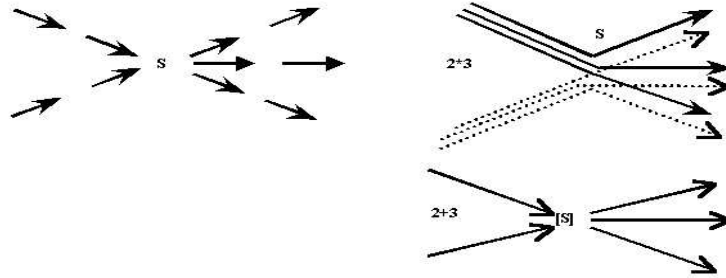
Figure 3.3: **(a)** System illustrating the combinatorial explosion. **(b)** 6 elementary flux modes if S is an internal metabolite. **(c)** 5 elementary flux modes if S is an external metabolite.

possibility to influence the whole system by the intervention on some branch-point metabolites was materialized in dietary purine restriction. Certainly a restriction might have effect only if the difference between the initial and the final configuration is big enough. Therefore this kind of management gives results only for those patients which normally ingest large quantities of purine-rich foods or drinks.

A way to mark that the intervention on a given metabolite is allowed is to set it as external. Switching an internal metabolite into an external one, the map of the elementary flux modes is also changed. If the switched metabolite is a branch-point metabolite (or hub-metabolite as [JTA$^+$00] call it), often the number of elementary flux modes decreases. This is because a combinatorial explosion is avoided [SKW$^+$02], a case presented in a small example in Fig. 3.3. Very interesting are the systems, whose topology allows recovery of the initial elementary flux modes from the elementary flux modes of the modified system through combinations methods, such as in the case of Fig. 3.3. This kind of systems permits a simplification during the study, but they do not lose information. There are, however, also systems in which the determination of elementary modes in the entire system by combination of the modes in the subsystems is difficult.

In the system under study, there is no conservation relation. This fact is explicable in a system which has as purpose to produce and decompose nucleotides. Usually in other systems (see Fig. 4.3) where ATP is only an energy source, $ATP + ADP = const.$

The "de novo"pathway is very expensive in comparison with so called "salvage" one. To produce one IMP, 6 ATP are spent. Therefore, most organisms rely mainly on the salvage pathways, which are able to convert more conveniently (with maximal 2 ATP and/or 2 NAD) the existing nucleotides in required nucleotides. Species not having "de novo" pathways themselves

have to use properly the "de novo" pathway of the host. This means that the chosen host has to be able to provide all the essential genetic precursors.

Table 3.2: Hierarchy of enzymes and the diseases that their deficiency caused.[*]

Enzymes	Number[**]	Diseases
PNPase	77	Lymphopenia and ISCD[***]
AMPase	64	Lymphopenia and ISCD[***]
PRPPs	64	Hyperuricemia
Xd2	56	Xanthinuria
HPRT	49	Hyperuricemia Lesch-Nyhan syndrome
ADA	41	Lymphopenia and immunodeficiency severe combined immunodeficiency disease
ADK	29	Contributes to neonatal hepatic steatosis
APRT	25	Crystaluria and the formation of urinary stones
XOR	20	Xanthinuria
GDA	19	High activity in acute and chronic hepatitis, cirrhosis of liver and C virus hepatitis
CNPDe	2	Human retinitis pigmentosa
KAD	1	Hemolytic anemia
NTDPase3	1	Thiopurine drugs toxicity
AIRs	1	Psychomotor regression
ATPase	1	Myopathy in human
GARs	1	Hyperuricemia in Down syndrome

[*] In our model, ADc, AICARt, AIRc, AIRc2, AIRs2, AmidoPRt, AM-Pda, ASS, FGAMs, GARt, GDPase, GMPr, GMPs, GUc, GUK, IMPc, IMPd, KPY, NDK3, SACAIRs are involved in one elementary flux mode. In the literature, we did not find any direct reference to a specific disease. For several such enzymes, only animal models were investigated.

[**] Number of elementary flux modes in which the enzyme is involved

[***] Immunodeficiency severe combined disease

Importantly, 65 elementary flux modes use R5P to produce nucleotides. One of them is "de novo" pathway and the rest are part of the salvage pathway. Only 20 modes can convert nucleotides without using R5P. Because R5P is an important product of the pentose phosphate pathway, problems occurring there propagate also in nucleotide metabolism.

HPRT and/or APRT are part of 64 elementary flux modes. Among the remaining elementary flux modes, only 5 are able to produce ATP and 4

of them have as end-product also urate. Obviously, a defect in HPRT and APRT will have grave consequences; this result is confirmed also by experimentally studies.

In Table 3.2, the hierarchy of most used enzymes is presented together with the diseases caused by the malfunctioning of some of them. For such very used enzymes, it is interesting to find out which are their products and then to search alternatives routes capable to yield these products. One can manage such a problem fully automatically, going twice through the set of elementary flux modes. First time, one has to record the elementary flux modes that do not use the enzyme under study and the end-products of the modes containing this enzyme. Second time, one has to examine the list of modes resulted in the first step and find out which of them could provide the end-products of the modes containing the specified enzyme. Also between these end-products it is recommendable to set an hierarchy. Unfortunately, this can not be done automatically. Strictly biochemical knowledge can decide which of these end-products are really needed for the good functioning of the organism. If a set of modes leading to such a very important end-product is empty, we can conclude that the enzyme under study is an irreplaceable enzyme. Defects in this enzyme cause without doubt severe diseases. If the set of modes is not empty, but is also not so rich, it might be that defects in expression of this enzyme cause diseases that are not so severe.

In the next section, we shall recall those diseases which were associated experimentally with the deficiencies of PNPase, PRPPs, HPRT3, ADA, APRT, and XOR, which are the enzymes with the highest involvement degree.

3.3   Impact of enzymes defects on health

Hyperuricemia (serum urate concentration exceeding 7.0 mg/dl in men and 6.0 mg/dl in women) and gout (urate crystal deposition disease) can be caused by inborn metabolic errors altering uric acid homeostasis. Three enzymes defects lead to hyperuricemia and gout [Bec00]. In glucose-6-phosphatase deficiency (glycogen storage disease called also Gierke's disease) both excessive uric acid production and impaired uric acid excretion occur. In severe and partial deficiencies of HPRT (number 5 in our hierarchy) and in super activity of 5-phosphoribosyl-1-pyrophosphate synthetase (PRPPs – number 2 in our hierarchy), purine and uric acid overproduction favour hyperuricemia and gout. Super activity of PRPPs can occur due to loss of feedback inhibition or increased affinity for R5P.

Xanthine oxidoreductase (XOR – number 9 in our hierarchy) catalyses the oxidation of hypoxanthine to xanthine and then to uric acid. Inherited

deficiency of this enzyme is very rare and usually asymptomatic or benign [RSL00]. It causes the precipitation of xanthine in the urinary tract or muscle. This may give rise to urolithiasis and muscle pain. Xanthinuria – the disease caused by XOR deficiency – is characterized by low uric acid and elevated xanthine in plasma and urine. The treatment consists in low-purine diet and high fluid intake.

An almost effective way to suppress excess uric acid production due to, for example, HPRT deficiency, consists in xanthine oxidase inhibition using allopuritol. However, this may have severe side-effects caused by xanthine accumulation. Among the 11 elementary flux modes containing XOR and not containing HPRT, 3 modes contain also PNPase5, another 3 contain also GDA and the remaining 5 modes contain Xd2. Each of these enzymes produces xanthine, which has to be further degraded by XOR to urate. If XOR is inhibited by allopuritol to avoid urate production and accumulation, also xanthine synthesis from hypoxanthine will be stopped, but not xanthine production by one of those above-mentioned enzymes. Since xanthine accumulation leads to xanthinuria, it is clear why the treatment with alloporitol often has side-effects.

Dietary purine restriction is not effective in the treatment of patients with normal dietary habits, but it may improve the state of those who normally ingest large quantities of organ-rich foods (liver, sweetbreads), beer or distilled spirits.

Lesch-Nyhan syndrome (LNS) is the most severe disorder caused by deficiencies of HPRT (number 5 in our hierarchy). LNS is always accompanied by hyperuricemia and gout [JF00]. Their damage degrees depend on the amount of residual enzyme activity. Unfortunately, the treatment with allopuritol is not able to cure the neurobehavioral disorders which attend the Lesch-Nyhan syndrome. To explain this aspect, an analysis on the processes going on in the brain has to be done. Certainly, several lesions caused at that level by HPRT deficiency are permanent. Until now, more than 200 mutations responsible for this disease were characterized. However, this aids only the development of rapid and convenient methods for diagnosis and prenatal testing. Effective treatment strategies for neurobehavioral features were not yet elucidated.

Unlike HPRT, adenine phosphoribosyltransferase (APRT- number 8 in our hierarchy) is not vital for the overall control of purine metabolism in humans [STKS00]. Its deficiency results in adenine oxidation to 2,8-dihydroxy-adenine (2,8-DHA). This process, catalyzed by two enzymes, is not shown in our scheme because it is not considered part of the nucleotide metabolism. 2,8-DHA is very insoluble and its accumulation in kidney leads to crystaluria and the formation of urinary stones. Up to now 18 mutations causing APRT

deficiency were identified. Although a dietary purine restriction showed only slight results, it was suggested that diet may be an important precipitating factor in the clinical expression of the defect. This assumption is supported by the fact that members of the communities that consume adenine-rich diets present more often the signs of the enzyme deficiency. Allopuritol therapy was used to control 2,8-DHA formation and a high fluid intake together with a low-purine/low-adenine diet were recommended.

Adenosine deaminase (ADA - number 6 in our hierarchy) and purine nucleoside phosphorylase (PNPase - number 1 in our hierarchy) have a great degree of involvement in the system. The diseases caused by deficiencies of these enzymes confirm the relevance of such an evaluation criterion. ADA and PNPase deficiencies impair lymphocyte differentiation, viability and function, resulting in lymphopenia and immunodeficiency [HM00]. Often, these patients lack both cell-mediated (T cell) and humoral (B cell) immunity, which lead to severe combined immunodeficiency disease (SCID).

Let us take a more close look in the mechanisms of ADA and PNPase. The elementary flux mode concept is also here suitable. When PNPase does not operate, only 21 out of 97 elementary flux modes are possible. One of these represents the "de novo" synthesis (mode 33 in Appendix C) – as already said, a very expensive mode - and the others are very simple modes (modes 1-20 in Appendix C), most of them being irreversible and responsible for conversion between the few existing nucleotides. Roughly speaking, a cell lacking PNPase does not dispose of the nucleotides necessary to synthesise the DNA for replication. [HM00] said that the lymphocytes are not able of differentiation, but we assess that they are also not able to replicate. Moreover, if the mechanism of cell apoptosis is still functioning, the lymphocytes number decreases, leading to lymphopenia. The whole organism is consequently not any more able to fight against infection agents, resulting in immunodeficiency. Also the modes producing urate are blocked. We can consider this fact as a measure to avoid the inopportune lost of nucleotide when they are, anyway, not sufficiently available. Unfortunately, in the case of severe PNPase deficiency, no cure is nowadays available.

Our model excludes the whole deoxymetabolism, focusing on, let us call it so, the central nucleotide metabolism. However, several experimental findings draw our attention also to deoxymetabolism. In the absence of ADA, deoxyadenosine is phosphorylated to yield levels of dATP that are 50-fold higher than normal (http://www.amg.gda.pl/~essppmm/ppd/ppd_pu_ada.html). High concentrations of dATP, especially in lymphocytes inhibit ribonucleotide reductase, thereby preventing other dNTPs from being produced. The net effect is to inhibit DNA synthesis. Due to inability to produce enough lymphocytes in response to antigenic challenge, SCID occurs. It would be of

interest to analyse also the deoxymetabolism considering the elementary flux modes. It is very likely to reach a theoretical conclusion in agreement with the experimental observations, but with less effort.

The treatments for ADA deficiency consist in bone marrow transplantation from human leukocyte antigen identical donor, which gives a complete or partial immune reconstitution, a T cell-depleted marrow transplantation from an haploidentical donor, which is less effective and also presents higher risks, or replacement therapy by intramuscular injection of bovine ADA modified by attachment of polyethylene glycol (PEG-ADA). This shows that a deficient enzyme could be also directly supplied. Usually, two risks have to be avoided. Firstly, if the enzyme is administrated orally, due to its big dimensions, it will certainly be cut by the digestive tract enzymes. Secondly, if it is administered intravenously, it might trigger the immunoresponse, so it could also be destroyed. It may be argued, however, that the second method can be successfully applied because, in the case of SCID, the immunoresponse should be almost null.

Adenylosuccinate lyase (4.3.2.2) deficiency (ADSL – enzyme participating in the only "de novo" elementary flux mode) is characterized by the appearance of succinylaminoimidazolecarboxamide riboside (SAICAriboside) and succinyladenoside (S-Ado) – the substrates of ADSL - in cerebrospinal fluid, urine and, to a much smaller extent, in plasma [VdJ00]. The main problem is not that the "de novo" pathway is blocked, but the accumulation of SAICAriboside and S-Ado seems to be toxic. It causes variable degrees of psychomotor retardation (also muscular wasting), often accompanied by epileptic seizures and/or autistic features. All 19 mutations identified up to now seem to lead to structural instability of the enzyme, without modifications of its kinetic properties. The treatment with oral supplements of adenine and allopuritol (the latter to avoid conversion of adenine into 2,8-DHA) showed no improvements. Very likely, the adenine was degraded before reaching where it was necessary.

In the medical literature, there is no direct reference to a disease caused by AMPda. Moreover, if in our system the status of IMP is changed from external to internal and the status of inosine is changed from internal to external, one can observe three parallel routes producing inosine (Inset of Fig. 3.2). The first mode contains AMPda and AMPase3, the second - ADK and ADA, and the third – AMPase and ADA. So inosine can be produced without using AMPda. It might be that the existence of this parallel route is enough to prevent a severe disease.

3.4 Medical approaches against proliferative and autoimmune diseases in the light of pathway analysis

A common feature of cancerous cells is that they grow and multiply rapidly. To avoid this phenomenon, a possibility is to find promptly reacting targets for chemotherapy in the purine or pyrimidine pathways of these aberrant cells. Disrupting of one of these pathways, the cell is deprived of the nucleotides necessary for its fast multiplication and this is often effective in treating proliferative diseases. For example, dihydroorotate dehydrogenase (DHODH - belonging to the pyrimidine pathway) has been found to be a target for the treatment of autoimmune diseases as rheumatoid arthritis as well as a target for immunosuppression during transplantation [FHF+99].
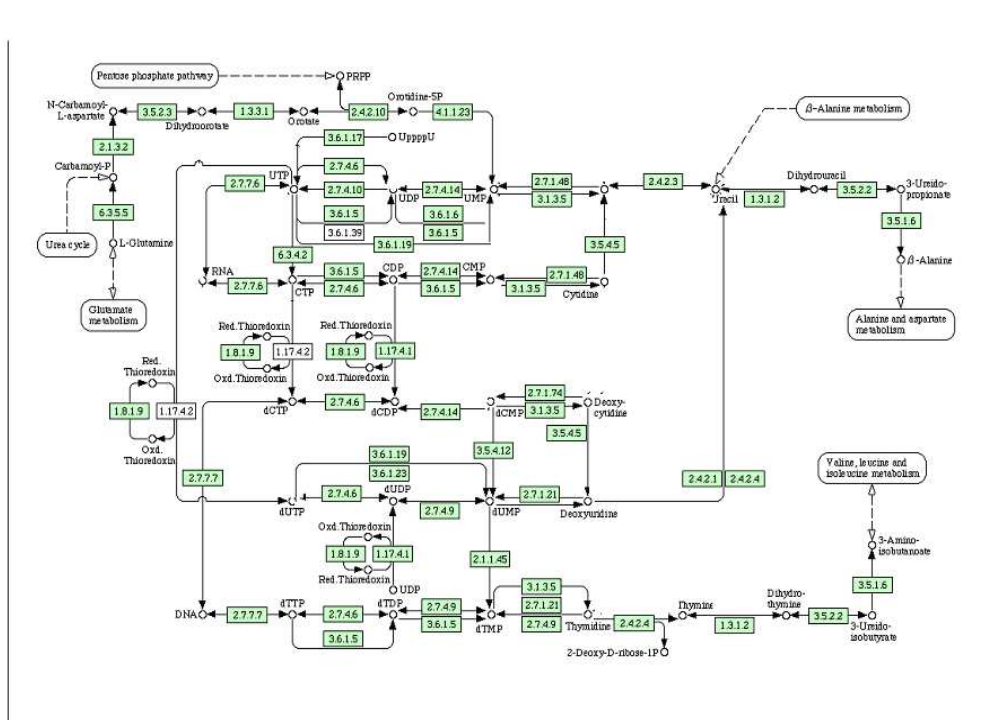

Figure 3.4: Pyrimidine metabolism extracted from KEGG. Cycles: metabolites; grey rectangles: E.C. numbers of the enzymes present in humans. Names and detailed reactions are given in Appendix D.

Let us consider the pyrimidine metabolism for humans as it is given in the database KEGG at

www.genome.ad.jp/dbget-bin/get_pathway?org_name=hsa&mapno=00240.

It is also depicted in Fig 3.4. Appendix D presents the names, abbreviations and E.C. numbers of the enzymes participating in this system and the reactions that these enzymes catalyse.,

The input file used by Metatool to calculate the convex basis, the enzymes subsets, the elementary flux modes, and the branch metabolites is given in Appendix E. PRPP, Gln, Glu, PPi, CO2, H, O2, H2O2, H2O, NH3, β-alanine, AisoB, ME4HF, DHF, R1P, dR1P, aspartate, dCTP, dTMP, dUMP and UMP are considered as external metabolites because they are either sources or sinks. Pi, ATP, ADP, AMP, NADP, NADPH, and NADPH2 are also treated as external due to their high degree of connectivity. On these conditions, we obtained 64 elementary flux modes showed in Appendix F.

DHODH is involved in all 5 elementary flux modes that stand for the "de novo" synthesis of β-alanine. Thus, knocking out DHODH forces the cell to use only the already existing nucleotides through its salvage pathways. There are another 30 elementary flux modes producing ß-alanine from dCTP. In KEGG, there is only one enzyme (MAD2L2 – E.C. number 2.7.7.7) producing dCTP from DNA. We omitted this enzyme in our system because cancer cells and lymphocytes need DNA for proliferation and do not, thus, degrade it. For young cancerous cell, DHODH is essential, because it is necessary to produce some pyrimidines to be then able to reuse them. But otherwise, from our analysis it can be deduced that it is not enough to inhibit DHODH. Also the alternative routes leading to β-alanine have to be interrupted. Because the number of substitutive routes is already high, one has to find an enzyme present in all elementary modes or the minimal combination of enzymes the inhibition of which stop the flux in all these modes. In this case, there are three enzymes present in all these modes: DPYD1, DPYS1, and UPB1 – they form also an enzyme subset. Knocking out any of them blocks each mode producing β-alanine. If, for various reasons it is difficult to provide drugs for the inhibition of each of them, there are still several interesting enzymes. Each enzyme from this enzyme subset S = {DHODH, -UMPS1, -CAD1, CAD3, UMPS2, CAD2} is essential for de novo modes. UP or NM3 are essential for all modes producing β-alanine. UMPK1 is contained in almost all modes producing β-alanine and those modes that do not use it, use UMPK2. As UMPK is an enzyme with low specificity, both functions can be blocked by inhibition. Therefore, inhibiting any of the enzymes belonging to the subset S or one of the enzymes UP, NMP or UMPK could be worthwhile for treating proliferative diseases.

[MWY+03] observed that thymidine kinase and thymidylate synthetase activities correlate positively with the stage and grade of renal cell carci-

noma. They suggested that thymidine kinase activity may be associated with the malignant potential of renal cell carcinoma and thymidine kinase and thymidylate synthase inhibitors may be a molecular therapeutic target for this disease.

The thymine nucleotide precursors for DNA synthesis depend on thymidylate synthase (TYMS). After more than 40 years, when TYMS was the primary target of 5-fluorouracil (5-FU), the search for TYMS inhibitors continues [DMS99]. The interest is straightened now to structural analogues to 5,10-methylenetetrahydrofolate, the second substrate of TYMS. The major problem of this class of inhibitors is that TS is not the only enzyme that they affect.

As in our model, the elementary flux modes in which TYMS acts represent the only source of AisoB, which is used further in valine, leucine and isoleucine metabolism. Non-functioning of TYMS leaves indeed the cell without its DNA precursors. Therefore, we can expect that TYMS can be an effective target in the fight against cancer.

For human cancer, the inhibition of the enzymes of the pyrimidine pathway seems to be accompanied by toxic side effects [CLW02]. It is difficult to predict how drugs such as methotrexate (MTX) and 6-mercaptopurine (6MP), acting on the pyrimidine pathway [MKK01], manipulate the cell, since they have multiple sites of action. Recently, inhibitors rationally designed from the knowledge of the catalytic mechanism, based upon the X-ray structure of the target enzyme, can become drugs with only one site of action. An example is VX-497, which inhibits IMPDH involved in the purine pathway. Therefore, this pathway may be a safer target for inhibition.

Biosynthesis of guanine nucleotides has been reported to be upregulated in tumour cells. Therefore, IMP dehydrogenase (IMPDH) – one of the rate limiting enzymes in guanine nucleotides synthesis, represents also an interesting target in the purine pathway for cancer chemotherapy. It was experimentally shown that, by blocking the conversion of IMP to XMP, IMPDH inhibitors lead to depletion of the guanylate (GMP, GDP, GTP and dGTP) pools. Both nucleoside inhibitors, such as ribavirin and tiazofurin, and non-nucleoside inhibitors, such as mychophenolic acid, showed antineoplastic, antiviral and immunosuppressive activity [FG99].

Indeed, there is only one elementary flux mode that can transform glutamine to GMP. It depends on XMP. Each other mode producing GMP uses GTP or GDP as reactant and vice-versa. We can conclude that XMP is essential for guanylate pools. But looking for alternative elementary flux modes capable to produce XMP, we found out another 24 elementary flux modes (given in Appendix F). All of them use R5P and some of them also need IMP. Importantly, they are very expensive, consuming relatively high

amounts of ADP, or ATP, and NAD. It might be that the cell avoids to use them and the decrease in guanylate pools was great enough to be observable. But it might also be that blocking simultaneously all the modes producing XMP leads to an even more effective result. Both PNPase2 and PRRPs and various functions of AMPase act in each such elementary flux mode. So, inhibiting any of these enzyme together with IMPDH stops the feeding of guanylate pools. But we must also consider the side-effects of such an action. As already presented in the previous section, the lack of PNPase and / or AMPase leads to severe combined immunodeficiency disease. PRPPs deficiency was associated with mental retardation, hypouricemia, megaloblastic changes in the bone marrow, and increased excretion of orotic acid in the urine [WNT$^+$74]. Therefore inhibition of any of these enzyme can be a dangerous solution. It is applicable only for a very short period of time to avoid that the conditions that trigger the disease are fulfilled.

It was suggested that glycinamide ribonucleotide transformylase (GARt) – an enzyme in the de novo purine nucleotide biosynthesis pathway - is a target for a series of compounds whose structures resemble that of tetrahydrofolate [SBS91]. Its inhibition is sufficient to induce the maturation of HL-60 leukaemia cells.

As GARt is an enzyme belonging to the only de novo elementary flux mode, its inhibition prevents the cell from producing IMP from R5P and forces it to only reuse the few available nucleotides.

In the purine pathway, PNP is involved in 77 out of 97 elementary flux modes. Also its inhibitors have the potential to suppress the T-cell response in T-cell proliferative diseases such as T-cell lymphoma and T-cell leukaemia, as well as in T-cell autoimmune diseases such as rheumatoid arthritis and lupus. This method of suppressing T-cells may also find application in organ transplantation. Despite the potential benefits of a potent inhibitor of PNP activity, no such therapy was tried up to now [PE02].

Most of the studies undertaken to find potential drugs were mainly experimental. Nevertheless, few others drew successfully the attention to the possibility to rationally design drugs based on the prediction given by pathway analysis. For example, [CCV$^+$00] showed that tranketolase inhibition in pentose phosphate pathway leads to the disappearance of several important elementary modes, whereas its activation increases the flux in the system and speed up the synthesis of ribose, which is incorporated in nucleic acids necessary for tumour growth, chemotherapy resistance, and proliferation. Moreover, this study resulted interestingly, in a fact that [Bor00] emphasised. Thiamine (vitamine B$_1$), administrated as fortifier, turn out to increase, in fact, the risk of cancer because it is an activator of tranketolase.

3.5   Similarity and difference between enzyme deficiencies and enzymes "hijacked" by viruses or parasites. Medical trials to cure them in the light of pathway analysis.

Some viral mechanisms are already known, others have only been hypothesised. Importantly, most viruses disturb the host cell nucleotides mechanism, to use it in their own interest. "Hijacking" by viruses is similar to enzyme deficiencies and knock-out experiments because the enzyme is no longer available to the host. [YG84] showed what is behind Reye's syndrome, following an influenza infection. The viral RNA polymerase combines with and activates liver host cell ornithine decarboxylase, which is no longer available to combine and activate host cell RNA polymerase and is prevented from participating in the urea cycle. Thereby, mitochondrial carbamoyl phosphate levels increase until carbamoyl phosphate passes from mitochondria into the cytosol where it is metabolised by the de novo pyrimidine synthesis pathway. Briefly, host cell RNA polymerase is inactivated while viral RNA polymerase has complete access to newly synthesised pyrimidines, which are used for viral replication.

One of the most dangerous viruses of the last century, having an extremely high rate of replication and rapid development of drug-resistant mutants, human immunodeficiency virus (HIV), depends on host dNTP synthesis, due to the lack of specific enzymes of nucleotides metabolism, such as ribonucleotide reductase, nucleoside kinases, and deoxyribonucleases. Thus, the restriction of host dNTP synthesis may represent a strategy for both inhibiting HIV replication and slowing the development of drugs-resistant mutants. For example, hydroxyurea, a known inhibitor of ribonucleotide reductase, increases the anti-HIV activities of the substances AZT, 2,3-dideoxycytidine (ddCyd) and 2,3-dideoxyinosine (ddIno) in primary cultured human peripheral blood mononuclear cells [GJCM95].

Hepatitis B virus (HBV), causing both acute and chronic hepatitis was long time treated only with alpha interferon. The trials to combat HBV confronted with its capacity to restart its replication when the treatment is finished. [Kor95] proposed combination therapies, which have the advantage to reduce the amount of drug required for efficacy, and, thus drugs toxicity, and to induce a more effective and rapid shutdown of virus replication. He suggested that combination of lamivudine (3TC), a promising nucleoside analogue, with either interferon or penciclovir significantly enhances the

antiviral effectiveness of these of agents against HBV replication.

In humans, herpes simplex viruses (HSV) are a common cause of infections that can be sometimes severe, prolonged and life threatening. Penciclovir (PCV), acyclovir (ACV) and ganciclovir (GCV) are drugs that were successfully used to suppress HSV. Their target is HSV-thymidine kinase. HSV replication is inhibited by the selective phosphorylation of these compounds by a virus encoded thymidine kinase. But at an ever increasing frequency, thymidine kinase-negative ACV resistant strains of HSV occur [KSYA96]. [ZHM$^+$02] proposed another antiviral agent, N-methanocarbathymidine (N-MCT). It is a thymidine analogue incorporating a pseudosugar. It profoundly inhibits the development of HSV infections, recording no cytotoxicity against uninfected cells.

As presented here, a virus adopts a double-effective strategy. It transforms the host cell into a virus replication factory and, simultaneous, weakens its defence system almost to its breakdown. Importantly, the virus has no metabolism on its own, being able to "slave" the host metabolism. Although the host organism weakens because it does not dispose of any resources for its maintenance, it is not possible to supply the hijacked enzymes or their products by drugs as in enzyme deficiency diseases because the virus uses them also. Therefore, the already applied strategies were trying to knockout enzymes necessary in virus replication. This replication can indeed be stopped, but the host keeps to be deprived of some essential enzymes. Its immune system continues to be down and the victim does not cope with normal external attacks. Thus, to prevail over the viruses, one has to target the few specific virus proteins.

In contrast to viruses, the parasites have at least a small part of a metabolism. Knocking out some enzymes, existing in both host and parasite, could be sometimes dangerous. But kinetic studies can reveal which enzyme are intensively used by parasite, but only sporadically by the host. Choosing only such enzymes, one can avoid the unfavourable secondary effects in host. Another valuable approach is to detect an essential enzyme of the parasite, applying pathway analysis that we proposed, to create a mutant of the parasite and to inoculate it as a vaccine, waiting then for the immune response that the mutant recognised as antigen triggers. Such a strategy, but starting rather from experimentally work, was successfully followed for *Toxoplasma gondii.*

T. gondii is one of the most widespread parasites of wild and domestic animals, provoking toxoplasmosis [CJ01]. It follows a very complex life cycle, having as intermediate hosts many and varied organisms and as definitive host – the domestic cat. Therefore, it can be regarded as an opportunistic parasite. Due to the close association between humans and cats, high levels

of infection with *T. gondii* occur in humans. Normally, toxoplasmosis has no visible symptoms, but it can also be lethal in immunocompromised patients and causes severe birth defects in newborns from primary infections during pregnancy. *Plasmodium falciparum* belongs to the same family of intracellular parasites and causes a virulent malaria.

Importantly, *T. gondii* could not rely on the pyrimidines provided by the host. Therefore it retained its ability to synthesise these essential genetic precursors. The same is valid for other members of its family. [FB02] managed to create a mutant *T. gondii* knocking-out one enzyme of its "de novo" pathway (uracil phosphoribosyltransferase). This mutant was not any more able to replicate itself. It lost its virulence; no symptom occurred in mice infected with a dose of mutant that would be lethal, if it contained wild-type *T. gondii*. Even immunocompromised mice did not present any warning sign. Essentially, it provided a high protection against the normal parasite, acting as a vaccine.

As this section illustrates, viruses and parasites attack especially the nucleotide metabolism to destroy the host immune defence and to use purines and pyrimidines of the host for viral and parasite replication. The faster a virus replicates, the faster it can develop drugs-resistant mutants. Applying their own strategy could be a method to prevent them from replication and even to kill them. Therefore, most treatments consist in inhibition of one or some important enzymes of the nucleotide metabolism which are vital for viral replication. Unfortunately, by blocking such enzymes, often not only the virus is hit, but also the host. Either we make drugs, which are directed at enzymes present in both host and virus, but capable to distinguish between them or we should aim only at virus-specific enzymes. Moreover, metabolic control analysis can help to assess whether an enzyme exerts high control in the virus and low control in the host. For all these strategies, a good understanding of virus mechanisms is required. And then, analysis of elementary flux modes existent in the viral system could help us to refine the methods.

Chapter 4

Modelling metabolic networks as Petri nets

4.1 Similarities between Petri net theory and traditional biochemical modelling

In graphical representations of Petri nets, circles are used for places, while rectangles stand for transitions (Fig. 4.1). The correspondence place – substance (in biochemistry often called metabolite) and transition – reaction/enzyme is obvious. Metabolic networks have a static level - the stoichiometry, and a dynamic one, characterized by fluxes. The stoichiometric coefficients indicate how many molecules of a substance have to react in order to produce how many molecules of product. The stoichiometric coefficients are described by the weights of arcs. Thus, the stoichiometry matrix containing these coefficients corresponds to the incidence matrix of a Petri net (see below).

A further object - the token- was introduced in order to describe the dynamics of a Petri net. It is denoted by a solid dot (•) inside the circles representing places. In ordinary Petri nets, the tokens do not represent specific information and are indistinguishable. They indicate the presence or absence of a condition, a signal, or a resource. In our case, the number of tokens in a place stands for how many molecules of that metabolite exist at a given moment. Alternatively, tokens may correspond to any predefined unit measuring the amount of substance, such as mole, millimole etc. However, this brings about that non-integer token numbers should be admitted. This leads to hibrid or continuous Petri nets, which are currently being developed [AD98, MDNM00]. Executable Petri nets models were proposed by [GKV01], as a first step to automatic creation and implementation of high-level Petri
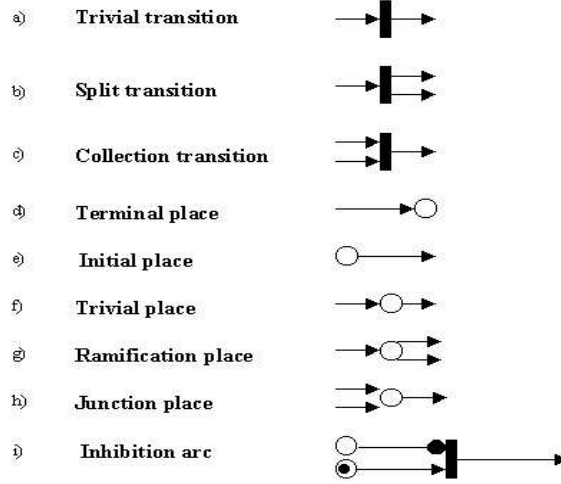
Figure 4.1: Components of a Petri net. Their counterparts in metabolic networks are as follows: **(a)** A→B (isomerization). **(b)** A→B + C. **(c)** A+B→C. **(d)** A product that is not further consumed. **(e)** A substrate that is not produced. **(f)** A metabolite produced and then consumed. **(g)** A metabolite produced in one reaction and then consumed in two or n reactions. **(h)** The situation opposite to **(g)**. **(i)** Inhibition phenomena.

Name	Notation	Definition
Preset of t	•t	{p $\mid$ p $\in$ P, pre(p,t) ≠ 0}
Postset of t	t•	{p $\mid$ p $\in$ P, post(t,p) ≠ 0}
Preset of p	•p	{t $\mid$ t $\in$ T, pre(p,t) ≠ 0}
Postset of p	p•	{t $\mid$ t $\in$ T, post(t,p) ≠ 0}

Table 4.1: Definition of the terms preset and postset.

net models [Jen97].

The tokens that exist in the system at a given time describe the state of the system. This is called marking, M(P) [Rei85, Sta90]. The system state changes when a transition fires. This can happen only if the transition is active/enabled, that means that every place from the input places set (Fig. 4.1 & Table 4.1) of the considered transition has at least as many tokens as the weight of the corresponding arc. The set of input and output places of a transition t is denoted by t and t, respectively. This set corresponds to the metabolites that act as reactants and products, respectively, in the reaction t. The new state is obtained by subtracting from each input place of the considered transition a number of tokens equal to the weight of the corre-

sponding arc and adding in each output place of the considered transition a number of tokens equal to the weight of the corresponding arc (Fig. 4.2 and Table 4.1). Formally, we can also speak about the input and output transitions set of a place p - p and p, respectively, which contain all the transitions which produce, respectively consume, the metabolite p.
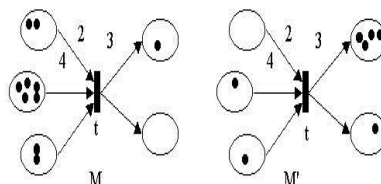

Figure 4.2: Marking and firing. M: P→N is called marking. For each place p∈P, M(p) represents the number of tokens which exist in p. M(p) gives the local state, while the vector M gives the state of the system and is called vector state. A transition is enabled/activated if $M(p) \geqslant pre(p,t) \ \forall p \in P$ and $K(p) \geqslant M(p)-pre(p,t)+post(t,p) \ \forall p \in P$. The mapping K: P→ N represents the maximal capacity of a place, if the number of tokens is limited. After the enabled transition fires, the new state of the system is M': P→ N, so that $M'(p) = M(p)-pre(p,t)+post(t,p) \ \forall p \in P$. In the example, the marking M'=[0, 1, 1, 4, 1] is obtained from the marking M=[2, 5, 2, 1, 0] after transition t fires. The formal description is: M [t > M'.

Considering the special graph description, it is useful to know between which places (P) and transitions (T) there exist arcs. For this purpose, two mappings describing weights were introduced (see also Table 4.1): **pre**: P×T→ **N** and **post**: T×P→ **N** (with **N** denoting the set of natural numbers). One can think about them also as matrices. The rows in **pre** correspond to places and the columns to transitions, while in the matrix **post**, the roles of rows and columns are transposed. The entries of these matrices have a nonzero value (equal to the weight of the arc), if an arc exists, and zero otherwise.

Further, the topological structure of a Petri net can be represented by an integer matrix, **C**, called an incidence or flow matrix. **C** is an n x m matrix whose m columns correspond to the transitions and n rows correspond to the places of the net. The following relation holds true: $\mathbf{C} = \mathbf{post}^T - \mathbf{pre}$. The mappings **pre** and **post** can be reconstructed from the matrix **C** in the following simple way:

$$\text{post}(t_j, p_i) = \max\{C_{ij}, 0\}, \ \text{pre}(p_i, t_j) = \min\{C_{ij}, 0\}.$$

It is worth finding whether another state can be reached from a given state. This is related with the property of reachability. In metabolic networks, we can search all possible subsequent states, knowing the initial state

of resources. Another interesting problem is to deduce all appropriate initial states from a desired later state.
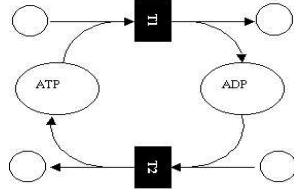

Figure 4.3: Simple example of capacity limitation in a metabolic system.

Places can hold an arbitrary number of tokens, or they can be restricted by a given number - capacitated places. In Fig. 4.3, the unnamed places are considered external (with inexhaustible numbers of tokens). The transitions T1 and T2 are activated if there is at least one token in the place ATP (the currency of metabolic energy), respectively ADP. If the initial marking is [c, c, c, c, 1, 0], transition T1 can fire and produce the marking [c, c, c, c, 0, 1]. Now, transition T2 is enabled. It fires and we obtain again the initial marking. In this example it is important that ATP + ADP = 1, independent of the system state. This is a conservation relation, which leads to boundedness of the capacity of all the internal places. Usually the places of biological systems are not considered to be limited because the limitation due to the finite size of living cells is not critical to most biochemical processes. There are only cases where the limitation comes from a conservation relation such as in the above case.

Another situation important in biological systems is the presence of inhibitors. The corresponding Petri net model can be extended by a special element, called inhibitory arc (Fig. 4.1i). The inhibitor is represented by a place. Of course, if there is a token at that place, the transition is not enabled, so it does not fire.

Note that the incidence matrix corresponds to the stoichiometric matrix [HS96] for metabolic networks if the Petri nets are pure. That means that the networks do not involve self-loops (Fig. 4.4), because self-loops cannot be represented in the incidence matrix: a coefficient 1 and a coefficient t1 cancel each other to yield zero in the matrix, thus losing track of the existence of the self-loop. Thus, we should identify the situations that produce self-loops and the way to treat them without losing the biological meaning. First, if the transition models only the reaction and the enzymes are considered as normal substrates, the problem can be avoided. There are algorithms for deleting / eliminating self-loops: For each loop, one introduces another place and another transition. If all weights equal unity, the number of tokens in

(a)   (b)
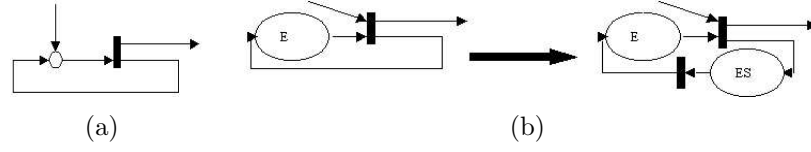
Figure 4.4: Self-loops. Left: the two types of self-loops. The conditions that a place and a transition are in a self-loop is: pre(p, t)·post(t, p) ≠ 0. In model **(b)**, the place marked by E represents an enzyme, which is regenerated after the reaction. The self-loop can be eliminated by decomposing the reaction in half-reactions (the two reactions that are depicted in the second part of **(b)** and considering the enzyme-substrate complex (ES). The new model represents a pure Petri net, which satisfies the relation: pre(p, t)·post(t, p) = 0 , ∀p∈P ∀t∈T. The number of tokens in the place ES is the difference between the capacity of the old place (the total amount of enzyme, we consider this equal to 1) and the number of tokens that place E contains already.

the new place is the difference between the capacity of the old place and the number of tokens that this old place contains already (Fig. 4.4(b)). By this construction, the new place corresponds to the enzyme-substrate complex and the so-called overall reaction (old transition) has been decomposed into half-reactions (the new transitions).
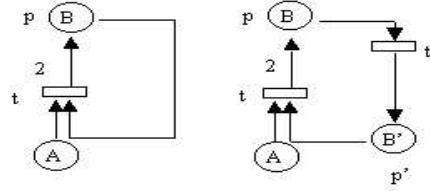
Figure 4.5: Petri net representation of autocatalysis.

This construction can be applied also when the arcs have a multiplicity larger than 1, but care should be taken that the new arcs might have also such multiplicities. If p' and t' stand for the new place, respectively for the new transition, the new arcs have to respect the formulae:

$$pre(p', t) \;=\; post(t', p') \;=\; pre(p, t') \;:= pre(p, t),$$

while the old arcs keep the same multiplicity. This situation can be nicely illustrated by the biochemical example of autocatalysis: A+B give 2B (Fig. 4.5). This reaction cannot simply be reduced to A gives B, because a small quantity of the product B is needed to start the reaction.

Second, a reversible reaction might be wrongly interpreted as a self-loop (Fig. 4.6). So, for each reversible reaction, we should consider only one flux
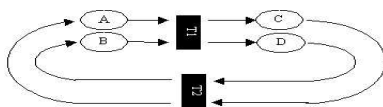
Figure 4.6: Treatment of reversible reactions. To model a reversible reaction A+B ↔ C+D, one usually introduces a transition for each direction.

direction and for the opposite one, we should introduce in the model another transition. So, the metabolic networks can be transformed into pure ones. Alternatively, one might think of newly defining reversible transitions. This has not, however, been dealt with so far in the literature.

A general feature of Petri nets consists in the possibility to be designed in different ways, called top-down and bottom-up. The first method supposes to start with a very generalized form of the system and then, to detail it as much as possible, until the basic units are reached. The second one starts with the "atoms", building modules which are then joint to model the real system. Sometimes it has many advantages to combine both of them. The importance of modularity is expressed by the ancient saying *"divide et impera"*.

4.2 Modelling of external metabolites

In metabolic networks, one needs to differentiate between internal and external metabolites. The internal metabolites are totally produced and then consumed in the given network, while the external metabolites represent sources or sinks [HS96]. Their amount is usually assumed to be constant, due to availability in large excess or well-tuned biological regulation. If one considers the given net as a part of a larger system, the external metabolites are a kind of boundary; or connection points with the remaining components. In this remaining part, pathways exist that have the function of producing or consuming these metabolites. An extension of the system in order to include those pathways is not useful, as the following example can illustrate. Glycolysis in humans (the well known pathway of sugar degradation [Str95]) contains a sequence of reactions that transforms glucose into pyruvate, producing ATP, which is the currency of energy in every organism. Glucose, pyruvate, ATP and also some other metabolites are usually considered "external" for this pathway. We might include, in the model, a reaction or pathway that consumes pyruvate, for example, for producing β-alanine. However, β-alanine then would be an external metabolite. The model needs to be delimited somewhere. In algebraic form the external metabolites can usually be identified also in the incidence matrix. Provided that each inter-

nal metabolite is both produced and consumed within the net, the external metabolites correspond to those rows in which all the coefficients have the same sign.

The modelling of external metabolites can be done in different ways. One of them is to fill all initial places with an inexhaustible number of tokens (modelled by infinity). For the sink places, one could allow them to accumulate tokens but has to take care in computing T-invariants (see below). If it is preferred to use finite token numbers for the initial places, one could redefine the firing rule for the transitions that have initial places in their preset or final places in their postset, in such a way to not "consume" the input-place tokens and to not "produce" final-place tokens,

$$M'(p) = M(p) \ \forall p \in I \cup F \tag{4.1}$$

where I is the set of initial places and F is the set of final/terminal places. Note that the firing rule for the internal metabolites is

$$M'(p) = M(p) - pre(p, t) + post(t, p) \ \forall p \in P \setminus (I \cup F). \tag{4.2}$$

Another possibility is to connect sink places with source places by additional transitions so that a circular flow occurs [HKS00]. However, it is difficult to find which places have exactly to be connected with each other, because such transitions could impose unrealistic constraints on the flow ratios. For example, one cannot regenerate carbon atoms from outgoing nitrogen atoms. A solution can be to use coloured Petri nets, in which different atom groups can be modelled by tokens of different colour. [Sta90] proposed, as another way of description, not to include the initial and final places in the net. Thus, the boundary is made up of transitions without presets or without postsets. The initial transitions do not need any tokens to fire. In the traditional modelling of metabolic networks, a similar description is indeed sometimes used for external metabolites that are of minor importance, such as inorganic phosphate, water, protons etc. However, applying this technique to all external metabolites has the drawback that they are not made explicit so that overall molar yields cannot be computed.

Here, we propose an alternative method. For each initial place, we add an arc feeding from the transition back to this place (Fig. 4.7(b)) and use the firing rule (4.2) both for internal and external metabolites. This guarantees that the number of tokens in the initial places remains unaltered. For each final place, we add an arc feeding from this place back to the transition producing it. To guarantee that the transition can always fire, at least one token should be put in the final place at the beginning. However, one should be aware that this generates self-loops, so that the Petri net is no longer

(a)

(b)
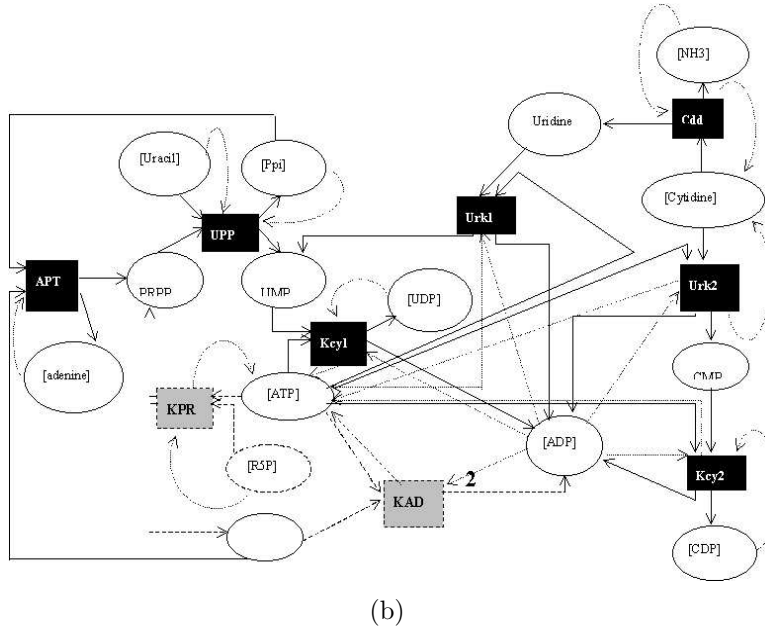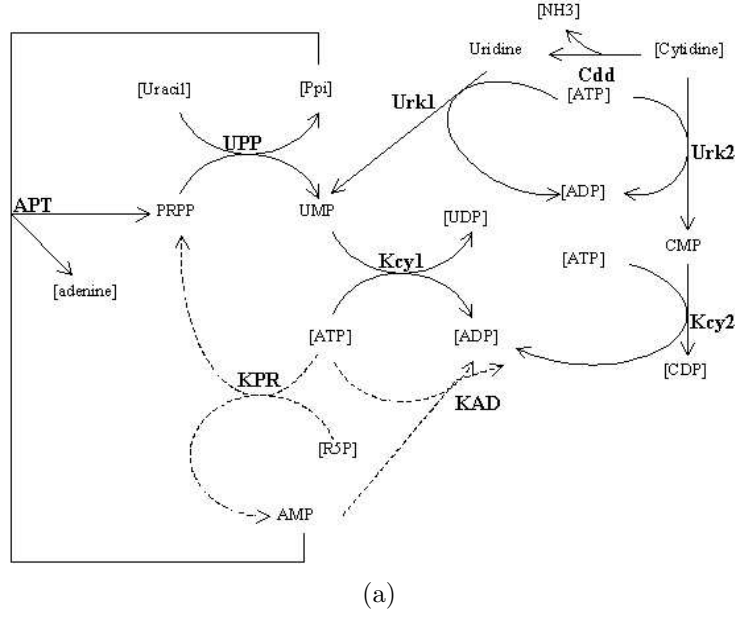
Figure 4.7: Part of nucleotide metabolism. The symbols stand for abbreviations of metabolites and enzymes usual in biochemistry. External metabolites are written in brackets. If the reactions indicated by dashed arrows are absent, the conservation relation ATP + ADP = const. holds. We do not consider R5P be part of the system in this case. **4.7(a)** Traditional biochemical representation. **4.7(b)** Petri net representation. The external metabolites were modelled with self-loops, depicted by dash-dotted arrows.

pure. Thus, as far as the external metabolites are concerned, the incidence matrix does not equal the stoichiometry matrix. This is no problem since the external metabolites are not usually included in the stoichiometry matrix [HS96].

4.3 Invariants in Petri nets

When studying a system, it is always appropriate to begin with the study of its structural invariants. They help in analysing the system's behaviour and checking its logical properties. The same is true for Petri nets describing biochemical networks because the structural invariants do not depend on kinetic enzyme parameters, which vary due to external influences and internal fluctuations. Basically, there are two types of invariants in Petri nets: P-invariants and T-invariants [Rei85, Sta90]. P-invariants (place invariants) are vectors, Y, with the property that multiplication of these vectors with any place marking reachable from a given initial marking yields the same result. If M_0 is the initial marking and M is some arbitrary marking, the relation $Y^T \cdot M = Y^T \cdot M_0$ describes a P-invariant and is called *relation of marking conservation*. Taking into account consecutive markings (that are obtained by firing of only one transition), it results that $Y^T \cdot col_t(\mathbf{C}) = 0$, for each transition t, where $\mathbf{C}$ is the incidence matrix. That means that, algebraically, these vectors are the solutions of the equation

$$Y^T \cdot \mathbf{C} = 0. \tag{4.3}$$

Invariants in Petri nets correspond to basic concepts in traditional biochemical modelling. In particular, P-invariants express conservation relations for metabolites, as becomes clear in the scheme shown in Fig. 4.3. This net has the P-invariant ATP+ADP=const. In general, equation 4.3 is known, as for metabolic systems, as the general form of conservation relations [Cla80, HS96]. In most cases, these relations express the conservation of atom groups [SH95b, SH91]. In the example in Fig. 4.3, the adenosine moiety is conserved.

In algebraic terms, invariants form a linear vector space. This implies that if I_1 and I_2 are invariants, also $c_1 I_1 + c_2 I_2$ with c_1, c_2 being real numbers, are invariants of the net [Rei85]. For example, if a biochemical net involves the P-invariants ATP+ADP=const. and NAD+NADH=const. [Str95], then also ATP+ADP+2 NAD+2 NADH=const. is a P-invariant. Normally, one chooses invariants with the smallest integer coefficients and tries to decompose the invariants into the minimal terms (such as ATP+ADP=const.). This leads to the concept of minimal invariants (see below).

In order that conservation relations reflect the conservation of atom groups (such as the adenosine moiety in ATP and ADP), the coefficients in these relations have to be non-negative. This leads to non-negative conservation relations [SH95b, SH91]. They correspond to semi-positive P-invariants in Petri nets [CS90]. If all substances are involved in non-negative conservation relations, the system is called conservative [ET89, HJ72]. This implies that a positive linear combination of all substance concentrations (token numbers in Petri nets) is constant in time,

$$\sum_i g_i z_i(\tau) = \mu, \ g_i > 0 \ \forall i \tag{4.4}$$

where τ denotes time. μ can be, for example, the number of some sort of atoms. In a closed system, that is, a system without external metabolites, there is always one relation of the type 4.4 in which μ represents total mass. In addition, there may be further relations of the type 4.4. If there is a positive linear combination that increases in time, the system is called superconservative [ET89]:

$$\sum_i g_i z_i(\tau) > \sum_i g_i z_i(\tau'), \ \tau > \tau', \ g_i > 0 \ \forall i \tag{4.5}$$

If the sum in Eq. 4.5 decreases in time until it reaches zero, the system is subconservative. The terms conservative, superconservative and subconservative have also been coined for Petri nets. The program INA developed by Starke and coworkers (www.informatik.hu-berlin.de/~starke/ina.html) determines whether a Petri net has one of these properties. Note that these three cases do not cover all networks. In fact, biochemical networks are usually open systems with a throughput of mass, as described by a flux between external metabolites. Therefore they may, depending on conditions, have a positive or negative mass balance, so that they usually belong to none of these classes.

P-invariants are useful in checking the property of mutual exclusion. Two transitions are called to be in mutual exclusion if there is no reachable marking that allows the two transitions to fire simultaneously. The first step is to identify the marking set that could characterise the simultaneous activation of the specified transitions. These markings are reachable only if they satisfy the conservation relation given by the P-invariants. For example, if four places need to have at least one token each to enable two transitions to fire simultaneously, while the conservation sum is three, mutual exclusion occurs. Importantly, such a case is irrelevant for metabolic networks analysis provided that the molecule numbers are large enough. Alternatively, if the

token numbers represent mole or millimole or the like, token numbers need not be integer, so that mutual exclusion is no problem either.

Often, two transitions are in mutual exclusion when they compete for the same input places set. If the tokens are indivisible and once a transition takes the existent tokens, the other transition cannot fire. If the tokens needed to reactivate the competing transitions are simultaneously regenerated for each conflict case, the net is called persistent. In this case, the two transitions do not deactivate each other. At first sight, some metabolic networks seam to be non-persistent because different enzyme reactions often compete for the same substance. However, in real metabolic nets, even if the quantity of the common resource is very small, the concurrent reactions share it, maybe in different percentage according to the various reactions rates. Until now, there are no techniques based on Petri nets that can model accurately this behaviour.

Biochemical networks often reach, after some initial transient, a stationary state. More concretely, this is the case when the kinetic properties of the networks are such that the stationary state is asymptotically stable, as is often the case [Cla80, HS96]. At steady state, the following equation holds:

$$\mathbf{C}V = 0 \qquad (4.6)$$

where V stands for the vector of net fluxes. They correspond to the flow of tokens per time in Petri nets. Special attention has to be paid to the involvement of external places in matrix $\mathbf{C}$. If we connect outputs with inputs by additional transitions as explained in the previous section or additional arcs are added to create self-loops next to initial and final places, $\mathbf{C}$ can contain the coefficients both for internal and external places. Otherwise, it should only contain the coefficients for the internal places in order for Eq. 4.6 to hold true.

A T-invariant (transition invariant) is a vector with the property that if each transition fires as many times as the value of the corresponding component of the vector indicates, the original marking is restored. Algebraically, these vectors are the solutions of Eq. 4.6. Therefore, T-invariants correspond to flux distributions in steady state.

As Petri nets usually involve irreversible transitions only, all components of a T-invariant must be non-negative. T-invariants with this property are called true T-invariants. Frequently, the net direction of all biochemical reactions in a network is known, for example, because they are irreversible or have a defined biochemical function. In this case, the orientation of reactions can be chosen in such a way that all (net) fluxes are non-negative. Then, only steady-state flux distributions corresponding to true T-invariants are

relevant.

Special interest in metabolic network analysis is paid to the elementary flux modes [SH94, SHWF02]. They stand for a minimal set of enzymes that could operate at steady state. That means that no other flux modes at steady state are proper subsets of the elementary flux modes. The choice of these elementary pathways is unique. For a better understanding of the behaviour of biochemical systems, they can be decomposed into such simplest relevant routes. This has been demonstrated for sugar cane metabolism [RB01] and bacterial metabolism [VDL02]. In Petri net theory, elementary modes have, as counterpart, the minimal T-invariants [SDF99]. The concept of elementary modes is, however, more general because reversible reactions are allowed. [CS90] developed an algorithm for computing minimal P-invariants. It can, by transposition of the incidence matrix, be used also for computing minimal T-invariants. The algorithm is based on row operations on the incidence matrix augmented with an identity matrix. In the course of calculation, care has to be taken to eliminate non-minimal and duplicate T-invariants. [CS90] propose two alternative tests to do so. A method for computing elementary flux modes based on convex analysis was proposed in [SH94, PSVNn$^+$99, SFD00]. Although the latter algorithm was developed with different goals ([CS90] did not deal with metabolic networks) and completely independently of Petri net theory, the two algorithms show some similarities. However, they differ in that elementary modes can involve reversible reactions. This is taken into account by partitioning the stoichiometry matrix into reversible and irreversible submatrices. Moreover, the test for eliminating non-minimal and duplicate T-invariants (elementary modes) is slightly different. For a more detailed comparison of the algorithms, see [SPM$^+$02].

The T-invariants are helpful in studying several properties of Petri nets, such as consistency. This property means that there exists an initial marking and a corresponding firing sequence that regenerates the initial state and contains each transition at least once. As can be seen in the system shown in Fig. 4.8 with either reactions 3 and 4 completely inhibited or reactions 5 and 6 completely inhibited, not every metabolic system is consistent according to this definition. However, reactions 5 and 6 in the former case and reactions 3 and 4 in the latter case are not covered by true T-invariants. If we consider only a subnet that is covered by true T-invariants, such as transitions t_1 and t_2 in the example, it is consistent. This is because, once the minimal T-invariants (elementary flux modes) are identified, appropriate initial markings enabling these invariants to operate can be linearly combined and a new initial marking is obtained. The system can fire each above-mentioned T-invariant consecutively (the necessary resources exist due to the "construction" of the initial marking), each transition is used at least once and the

initial marking is always regenerated.
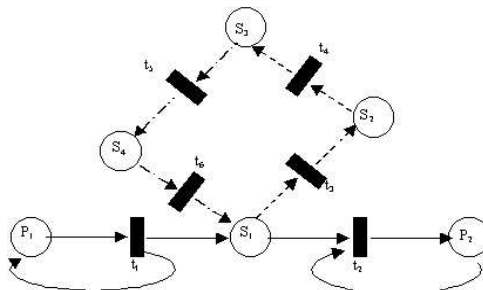
Figure 4.8: Traps and deadlocks. Two situations are considered: either transitions t_3 and t_4 are operative (dashed arcs) or the transitions t_5 and t_6 (dash-dotted arcs). P_1 and P_2, external metabolites; S_i, internal metabolites.

A further property studied for Petri nets, reversibility, means that for every marking M that can be reached from M_0, M_0 can also be reached from M. It holds for metabolic networks, if some constraints are fulfilled. One constraint is that the metabolic network is covered by true T-invariants. The second constraint is that all external metabolites have enough tokens to operate all true T-invariants. The arguments read as follows: Let us denote the number of times the transitions t_i have to fire in order to reach a marking M from M_0, by w_i. The numbers w_i are gathered in a vector W. Note that W need not fulfil Eq. 4.6. As the net is covered by true T-invariants, we can find a vector V that does satisfy Eq. 4.6 and a sufficiently large natural number λ such that $\lambda V - W$ involves positive components only. This vector indicates how many times the transitions need to be fired to reach the initial marking again.

4.4 Siphons, traps, deadlocks and liveness

In Petri nets, special sets of places can be identified, for example, siphons, called also structural deadlocks, and traps [Rei85]. A siphon is a set of places that – once it is unmarked – remains so. A trap is a set of places that – once it is sufficiently marked – can never lose all its tokens. (It can happen that, if only some places of the trap are marked with a number of tokens smaller than a certain limit, the trap may lose all its tokens.) Clearly, any semi-positive P-invariant implies a trap because the total number of tokens is constant and can, hence, not reach zero. Moreover, superconservative subnets form traps, while subconservative subnets form siphons. The algorithms calculating the siphons and traps [Sch96, TYW96, YTW96, YW99] in nets with specific

properties are based on the following alternative definitions: A siphon is a set of places having the property that its input transitions set is contained in its output transitions set. A trap is a set of places for which its output transitions set is contained in its input transitions set. A Petri net N, having m_0 as initial marking, is said to be deadlock-free if for any reachable marking m, there is an enabled transition. A Petri net N, characterized by a current marking m is in deadlock if no transition is enabled to fire at marking m. Preventing deadlocks in an efficient way represents an intensely researched field [Kem93, Var93, MA98, CX97, HJXC01, IMA00].

Traps and structural deadlocks are interesting for biochemical modelling. Many biochemical networks have the function to produce storage substances in certain periods and consume these substances in other periods. For example, the potato plant produces starch and accumulates it in the potato tubers during growth, while starch is consumed after the tubers are deposited after the harvest. The starch and several of its precursors then form traps in the reaction net during growth, while starch and possible intermediates of degradation form siphons after the harvest. Consider the simple reaction system shown in Fig. 4.8. Transition t_1 is always activated while t_2 only fires if at least one token exists in S_1. Importantly, we consider t_3 and t_4 to be inoperative if t_5 and t_6 are operative in the system and *vice versa*. For example, the system could describe the production and degradation of starch. The internal metabolites then would be: S_1, glucose-1-phosphate, S_2, UDP-glucose, S_3, starch [Str95]. In the starch example, it is not necessary to consider an intermediate S_4, while for other storage metabolites, it may be. In most cells containing starch, either the branch producing starch or the branch degrading it is functional. This is realized by complete inhibition of the appropriate enzymes. It can be easily observed that S_2 and S_3 form a trap when reactions 3 and 4 are operative. Once a token arrives in S_3, no transition able to fire exists in the system to consume this token, so it remains there independently of the later evolution of the system. In the other case, once the last tokens were extracted from S_3 and S_4, no transition able to generate a new token in these places exists, so they remain empty. This means that S_3 and S_4 form a siphon.

Current computer programs for simulating metabolic networks deal only partially with siphons and traps. For example, the program GEPASI developed by [Men97] (http://gepasi.dbs.aber.ac.uk/softw/gepasi.html) detects all reactions that are at equilibrium in any steady state. For the example system shown in Fig. 4.8 with transitions 5 and 6 blocked, GEPASI would detect reactions 3 and 4 to have this status. Here, these reactions 3 and 4 are irreversible, so that no steady state can be reached. However, if the reactions were reversible, they would indeed attain thermodynamic

equilibrium in any steady state. The Program METATOOL [PSVNn$^+$99]; (http://www.bioinf.mdc-berlin.de/projects/metabolic/metatool/) indicates metabolite S_3 as a "non-balanced internal metabolite", while it does not say anything about metabolites S_2 or S_4. It is worth including, in future refinements of simulation packages, routines for detecting all metabolites involved in traps or siphons.

Another important concept is liveness [Rei85]. A transition t is said to be live if, for any marking m reachable from m_0, there is a marking m' reachable from m, such that t is enabled by m'. A transition t is dead at marking m if no marking m' reachable from m enables t. A Petri net $(N; m_0)$ is said to be live if every transition is live.

Importantly, deadlock-freeness and liveness are two different notions. Liveness means that all the system's transitions may be repeated infinitely often, while deadlock-freeness only implies that at least a subset of these transitions may be repeated, but not necessarily all of these. It is interesting how the concepts of liveness, deadlocks, siphons and traps are connected with each other. A net satisfies the deadlock-trap property if each non-empty siphon includes a trap and the maximal trap in each minimal deadlock is sufficiently marked. In this case, no dead marking is reachable. So, the net is deadlock-free. A further special class of Petri nets is made up of the free-choice nets. In these nets, each place has at most one output transition or the input places of the output transitions of P consist only of P for any place P belonging to the net. A free-choice net is live if and nly if every non-empty siphon includes an initially marked trap. This property is also known as Commoner's theorem [Com72]. Siphon and trap are dual notions. A siphon in a Petri net N is a trap of the net N' obtained by reversing direction of all edges of N. Therefore, the properties satisfied by siphons have counterparts for traps.

Liveness and deadlock-freeness are structural properties, in which the initial marking plays, however, an essential role. An example for a net that is in deadlock is the above example A + B gives 2 B (Fig. 4.5) if the initial token number of B is zero. The input and output transition sets of B coincide. Therefore, {B} is simultaneously siphon and trap. ●{B●}={A,B}≠{B}, but $|●A| = |\{t\}|=1$ and $|●B| = |\{t\}|=1$, so "A + B gives 2 B" is a free choice net. Due to Commoner' theorem, this net can not be live if the token number of A is not infinite (if A is not what we called external metabolite in section 4.2) and trap {B} included in siphon {B} has not at least one token. Without tokens in B, this autocatalytic reaction then does not start proceeding. In chemistry, such a situation is known as false equilibrium [Oth81]. A larger example is glycolysis, which requires 2 moles of ATP in its upper part and produces 4 moles of ATP in its lower part. If no ATP is present at the

beginning, the glycolytic pathway cannot proceed. Therefore, this pathway has been said to have a turbo design [TWvDW98]. A test for liveness and deadlock-freeness of the net can thus help us decide whether the metabolic system can attain a situation where it is blocked. The detection of siphons and traps is instrumental for this purpose.

4.5   Evaluating the role of TPI in T. brucei metabolism by detecting siphons and traps

T. brucei is a unicellular, extracellular, eukaryotic parasite of the blood and tissue fluids of mammals. It is transmitted by tsetse flies and causes sleeping sickness in humans. The infections are lethal unless treated, but the few existing drugs have severe side-effects. Many studies (e.g. [BWtK$^+$99, HEB$^+$01] were focused on the carbon and free energy metabolism of this organism, which depends entirely on glycolysis. Accordingly to Scheme 1 in [HEB$^+$01], glucose is imported into the glycosome and then converted into F-1,6-P. The two consecutive enzymes (HXK and PFK) are contracted in one step which consumes two ATP and produces two ADP. ALD converts F-1,6-P into DHAP and GA3P. These two substances are isomerised into each other by a reversible enzyme, TPI (triose-phosphate isomerase). DHAP is transformed into Gly3P by GPDH, with consumption of NADH and production of NAD$^+$. GAPDH1 uses NAD$^+$ to transform GA3P into BPGA and NADH. Gly3P can be either converted into glycerol by GLYK with consumption of ADP and production of ATP, or transported into the cytosol, where GPO oxidizes it to DHAP and H$_2$O, DHAP being transported back to the glycosome. PGK uses an ADP molecule to convert BPGA in 3-PGA and ATP. 3-PGA is transported into the cytosol and converted, via 2-PGA, into PEP, which gives pyruvate and ATP, consuming one ADP. A similar system was treated by [OBK$^+$02]. They maximized the yield of glycerol in *Saccharomyces cerevisiae* using metabolic engineering.

At the beginning, [HEB$^+$01] supposed that glycolysis could proceed without TPI, producing glycerol and pyruvate in the same amount. Contradicted by the reality, they continued to study the case and built a kinetic model to explain the unexpected result of all system fluxes (PYR, GLYCEROL) decrease. We give now a structural explanation, ignoring, thus, the kinetics. In Fig. 4.9, the Petri net model corresponding to Scheme 1 in [HEB$^+$01] is given. It should be noted that this network is not a free-choice net because, for example, $|\text{GLY3P}\bullet| = |\{t2, \text{GLYK}\}|=2$ and $\bullet\{\text{GLY3P}\bullet\}=\{\text{GLY3P, ADP}\}\neq\{\text{GLY3P}\}$. The first property is known in

Petri net theory as conflict, because two transitions {t2, GLYK} compete for the same resources (the tokens from place B). But in metabolic networks, the token number is large enough that the transitions in competition will simply "agree" on the tokens distribution depending on their reaction rate. Taking this aspect into account, we do not need to know the reaction rates, but we only assume that the flux through transition t_2 in Fig. 4.9 is always greater than zero. Another important knowledge that we use is that ALD is reversible and therefore inhibited by its products [Str95].

Marking Firing sequences	ATP	ADP	Glu	F16P	DHAPg	DHAPc	GA3P	NADH	NAD⁺	GLY3Pg	GLY3Pc	BPGA	3PGA	2PGA	PEP	Pyr	Gly	
	2	0	∞	0	0	0	0	1	0	0	0	0	0	0	0	0	0	m_0
e_1	0	2	∞	1	0	0	0	1	0	0	0	0	0	0	0	0	0	
ALD	0	2	∞	0	1	0	1	1	0	0	0	0	0	0	0	0	0	
GPDH	0	2	∞	0	0	0	1	0	1	1	0	0	0	0	0	0	0	
GAPDH	0	2	∞	0	0	0	0	1	0	1	0	1	0	0	0	0	0	
PGK	1	1	∞	0	0	0	0	1	0	1	0	0	1	0	0	0	0	
t_3	1	1	∞	0	0	0	0	1	0	1	0	0	0	1	0	0	0	
e_2	1	1	∞	0	0	0	0	1	0	1	0	0	0	0	1	0	0	
e_3	1	1	∞	0	0	0	0	1	0	1	0	0	0	0	0	1	0	
t_2	1	1	∞	0	0	0	0	1	0	0	1	0	0	0	0	1	0	
GPO	1	1	∞	0	0	1	0	1	0	0	0	0	0	0	0	1	0	
t_1	1	1	∞	0	1	0	0	1	0	0	0	0	0	0	0	1	0	
GPDH	1	1	∞	0	0	0	0	0	1	1	0	0	0	0	0	1	0	
t_2	1	1	∞	0	0	0	0	0	1	0	1	0	0	0	0	1	0	
GPO	1	1	∞	0	0	1	0	0	1	0	0	0	0	0	0	1	0	
t_1	1	1	∞	0	1	0	0	0	1	0	0	0	0	0	0	1	0	dead marking

Table 4.2: Markings obtained during a firing sequence leading to a dead marking in the energy metabolism of *T. brucei*.

Let us consider the case when TPI is knocked out. T1={NADH, NAD⁺} forms a siphon and a trap at the same time. Its input transitions set {GPDH, GAPDH} coincides with its output transitions. This means that once this set of places is sufficiently marked, it keeps its tokens. Moreover, {NADH, NADH+} forms also a P-invariant, their tokens sum remaining constant during the whole process. Another trap (T2) consists of {DHAPc, DHAPg, GLY3Pc, GLY3Pg, Gly} because its input transitions set {ALD, GPDH, GPO, GLYK, t_1, t_2} includes its output transitions set {GPDH,

GPO, GLYK, t_1, t_2}. If the flux through t_2 were equal to zero, Gly would accumulate, but because the flux through t_2 is greater than zero, GLY3P is partially transformed back into DHAPg. Let us start proceeding with the marking m_0. Following the firing sequence {e_3, ALD, GPDH, GAPDH, t_2, GPO, t_1, GPDH, t_2, GPO, t_1, PGK, t_3, e_1, e2} as Table 4.2 illustrates, the network reaches a dead marking m_1, because DHAP accumulates - no NADH being available for further converting it. This is because GPO is draining the flux, consuming NADH faster than GAPDH can produce it. This continues until the product inhibition of ALD is so strong that ALD ceases operating. Therefore, the whole system is dead, no more Gly and Pyr being produced. For deriving this result, it is informative that after deleting TPI, the three transitions t_1, t_2 and GPO are not involved in any elementary mode (minimal T-invariants) anymore.

The importance of TPI can be seen if its corresponding transitions are added in the model. TPI being a reversible enzyme, two transitions, one acting forwards and one – backwards, have to be added. Of course, the T-invariant {TPI1, TPI2} has to be ignored, having no biological significance. In this new context, T2 is not any more a trap. Another minimal trap occurs in T3 = {DHAPC, DHAPg, GLY3Pc, GLY3Pg, GLY, GA3P, BPGA, 3PGK, 2PGK, PEP, PYR} corresponding to the accumulation of GLY and PYR. Whenever DHAPg tends to accumulate, due to NADH lack, TPI1 converts a part of DHAPg tokens into GA3P. Due to the sufficient amount of NAD^+, GAPDH fires with production of the necessary NADH, which gives GPDH the possibility to fire. Also if NAD^+ is deficient, but GA3P is sufficient, TPI2 converts GA3P into DHAPg, GPDH fires and produces the required NAD^+.

Taking into account almost only structural properties of the given network, especially the presence of traps, we could evaluate the role of TPI in glycolysis and glycerol production. In the next section an example taken from nucleotide metabolism will be used to illustrate the notions presented above. The program INA (www.informatik.hu-berlin.de/∼starke/ina.html) is utilized to facilitate the calculations.

4.6 Petri net properties on a system extracted from nucleoside metabolism

Let us now consider the biochemical system depicted in Fig. 4.7(a). It represents part of nucleotide metabolism, as it occurs, for example, in human liver [Str95]. We have translated it in terms of a Petri net (Fig. 4.7(b)) and then, analysed it using the program INA

Figure 4.9: Petri net representation of the glycolysis metabolism of Trypanosoma brucei T_1={NADH, NAD+} is simultaniously a P-invariant, a trap and a siphon. In absence of TPI_1 and TPI_2, T_2={DHAPg, DHAPc, GLY3Pg, GLY3Pc, Gly} is a trap. If TPI_1 and TPI_2 act, T2 is not a trap any more, but T_3={DHAPc, DHAPg, GLY3Pc, GLY3Pg, GLY, GA3P, BPGA, 3PGK, 2PGK, PEP, PYR} forms a trap. We assume that the flux through t_2 is greater than 0 and that there is a threshold of DHAPg, above which ALD is feedback inhibited.

For modelling the external metabolites, we have chosen to introduce a self-loop for each source and each sink and to keep the same firing rule (4.2) independently of the metabolite's type. If we do not impose capacities for the internal metabolites, the net is unbounded because uridine, for example, can accumulate more and more tokens if Cdd keeps firing while Urk1 is not. Accordingly, the reachability tree is infinite. As INA has reported, the net is strongly connected, not pure (which is obviously due to the introduced self-loops), and not (sub-)conservative. There is no P-invariant, except the external metabolites on their own. Again due to the self-loops next to the external metabolites, the number of tokens in each of these places remains constant.

INA reports four minimal semi-positive T-invariants. We give them here by indicating the transitions with non-negative components in the vectors representing these T-invariants:

1. Urk2, Kcy2

2. Cdd, Urk1, Kcy1

3. 2 Kcy1, KPR, 2 UPP, APT

4. Kcy1, KAD, KPR, UPP.

Note that some transitions (such as Kcy1 and UPP in the third invariant) have to fire twice, but not necessarily successively. One can see that firing all the activated transitions that belong to an invariant regenerates the initial marking. As each enzyme occurs in at least one minimal T-invariant, the net is covered by these invariants. Therefore, the net is persistent and live. For simplicity's sake, although in biological organisms the reactions KAD and KPR are reversible, we considered them irreversible. If they are treated as reversible, care has to be taken that extra, irrelevant T-invariants, containing only KAD and KAD', and KPR and KPR' respectively, result (where the primed symbols denote the reverse reactions). They have to be discarded.

The biochemical meaning of the minimal T-invariants can be explained as follows: Eq. (4.1) production of cytidine-diphosphate (CDP) from cytidine, Eq. (4.2) production of uridine-diphosphate (UDP) from cytidine, (3, 4) two invariants producing uridine-diphosphate (UDP) from uracil in different ways. In the invariant (3), one mole of adenine per two moles of UDP produced is formed as a by-product. This is because ATP is used as a

source of the ribose moiety, which is necessary for forming UDP from uracil. Note that this invariant is not easy to determine by inspection. Moreover, it can be seen that the molar yield with respect to ATP is different for the pathways (3) and (4). While invariant (3) consumes 3 moles of ATP per mole of UDP produced, invariant (4) uses 3 moles of ATP per two moles of UDP [SPM+02]. All of these T-invariants correspond to the so-called salvage pathways, which serve to save nucleotides from leaving the cell and redirect them to nucleotide phosphates [Str95].

Let us now assume that ATP and ADP are internal metabolites and that the two enzymes KPR and KAD are not expressed in a certain cell type. If we modify the network by eliminating the transitions that stand for these enzymes (and also the arcs that connect them with their neighbouring places), we do obtain a P-invariant. It can be translated in terms of a conservation relation: ATP + ADP = const. The constant would be 2 if we define the initial token numbers of ATP and ADP to be 1 each. With any conservation sum less than four, the four remaining transitions consuming ATP (Urk1, Urk2, Kcy1, and Kcy2) are in mutual exclusion. Since the net does not include transitions producing ATP or (in the second case) AMP, these substances are eventually running out, so that the places standing for ATP and AMP are siphons, while in the complete net, there is neither a trap nor a deadlock. To maintain a steady state, nucleotide metabolism requires permanent production of ATP, for example, by glycolysis.

Chapter 5

Routes in signalling maps

5.1 Formal description

5.1.1 Graph-theoretical description of monomolecular effects and reactions

The simplest signalling maps might contain only statements such as "A activates B"or "A inhibits B". Monomolecular reactions such as "A is converted into B"are another simple type of interaction. For networks only consisting of such effects and reactions, a proper modelling strategy can be based on labelled directed graphs.

When there are one activated and one or several inactivated forms of the same substance, we consider, in our model, only the activated one. Let us consider a linear signalling cascade, as that described in [KHWB97] and [HNR02]. $\tilde{X}_i$ stands for the dephosphorylated (inactive) form of the ith kinase, while x_i stands for the phosphorylated (active) form of the ith kinase. Assuming that the concentration of each kinase-substrate complex is small compared with the total concentration of the reaction partners, the concentration of each activated kinase as a function of time, $X_i(t)$, is given by the solution of the family of differential equations:

$$\frac{dX_i}{dt} = \tilde{\alpha}_i X_{i-1} \tilde{X}_i - \beta_i X_i \tag{5.1}$$

where $\tilde{\alpha}_i$ is the second order rate constant for phosphorylation by the ith kinase and β_i is the rate constant for dephosphorylation by the ith phosphatase [HNR02]. As the total amount of the active and inactive forms is constant, $C_i = \tilde{X}_i + X_i$, equation (5.1) becomes:

$$\frac{dX_i}{dt} = \tilde{\alpha}_i X_{i-1}(C_i - X_i) - \beta_i X_i \tag{5.2}$$

So, this equation can be written in terms of the concentration of the activated form of each substances only. Accordingly, we need to consider only the activated form in our model. Analogous considerations apply if there are more then two forms, provided that only one of them is active.

In the graph, the substances are represented by nodes. The activations or inhibitions correspond to arrows labelled by +1 or –1. If, for a given ordered pair (A,B), it is not known in which way (positive or negative) A influences B, the label will be 0. To find all the targets that are influenced by a given initial factor (F), it will be sufficient to extract the tree having F as root. Its leaves will be the searched targets. Going through nodes, from root to each leaf, the routes will be highlighted. If on a chosen route all the labels are different from zero, the sign of their product will provide us the information whether F has a positive or negative influence on the corresponding leaf. If one is interested in finding all factors that have an effect on a given target (T), the procedure is the same, with the exception that the root has to be T and the directionality of the arcs should be inverted when extracting the tree.

//input file for the theoretical example in figure 5.1(b)
#A
$F1/*F1$ is an initial factor $*/->I1$ // $F1$ activates $I1$
$F1->S5$ // $F1$ activates $S5$
$F2->S2$ // $F2$ activates $S2$
$F3->S4$
$S2->S3$
$S3->S7$ /* $S7$ is an intermediary */
$S4->S7$
$S1->S5$
$S5->T1$
$S5->T3$
$S6->T2$
$S8->S6$
$S7->S8$
$S7->T3$

#I
$I6->I3$ // $I6$ inhibits $I3$

Table 5.1: Input file for the theoretical example in 5.1(b)

In the algorithm, which we will describe below, the classification between initial factors, intermediaries and targets is made based on the number of

incoming and outgoing arrows. Thus, they need not be declared. However, sometimes ubiquitous substances such as H_2O or P_i could erroneously appear as factors or targets if they are only consumed or only produced, respectively. In these cases, we can just ignore them.

Some definitions are required: We say that on a path, a **node** is **balanced** if it has exactly one incoming arrow and exactly one outgoing arrow (belonging to the path). Such a node is called an intermediary. A **balanced route** has an initial factor F as initial node, a target as final node and between them, a succession (path) of arrows with balanced nodes in between.

As an example we consider the system shown in figure 5.1. We depicted the factors, intermediaries and targets using circles or ellipses. Each of them carries a label indicating its name, $A_1 - A_{13}$ - for activations, I_1 - for inhibition and R_1 – for reaction. We denote the different types of species as follows: $F_1 - F_3$ – initial factors, $S_1 - S_7$ – intermediaries and $T_1 - T_7$ – targets. However, as mentioned above, this specification is not necessary in the input file, which reads as shown in table 5.1. The algorithm makes this classification based on incoming and outgoing arrows. In the input file to the program SigNetRouter, activations, inhibitions and reactions are separated in different regions, which are started with the identifiers **#A**, **#I,** respectively **#R**. Even in such a small example, there are some routes and a cycle to be found (see figure 5.1(b)). They are more easily represented in tree form. In Fig. 5.1(c), the following sequence is a balanced route: the initial factor F_1 activates the intermediary S_1, S_1 activates S_5, and S_5 activates target T_1. Let us call it Route 1.

An **incomplete route** has an initial factor F as initial node, an intermediary L as final node and between them, a succession of balanced intermediaries, such that there is an intermediary S between F and L and an arrow leading from L to S. To maintain the tree structure, this arrow (which closes the cycle) should not be considered. This is the reason to introduce the concept of incomplete route. In figure 5.1(c), the following sequence is an incomplete route: $F_3 \rightarrow S_4 \rightarrow S_7 \rightarrow S_8 \rightarrow S_6 \rightarrow S_3$. Let us call it Route 2. A **cycle** is a succession of balanced intermediaries such that each node can be an initial node and a final node at the same time. In figures 5.1(c), the sequence: $S_7 \rightarrow S_8 \rightarrow S_6 \rightarrow S_3 \rightarrow S_7$ is a cycle. Let us call it Cycle 1. The cycles affect the routes that contain a part of them. In our example, Cycle 1 affects Route 2. If cycles exist in the network, at least one leaf of the tree extracted from the network can not be a target, but an intermediary which is the start node for an arrow, whose end node is on the route between F and this intermediary-leaf. In figures 5.1(c)-5.1(d), we depict the arrow that is closing the cycle with a dotted line. It has an opposite direction than the other arrows in the tree. It is not a part of the tree, since trees are graphs

(a)   (b)

(c)   (d)
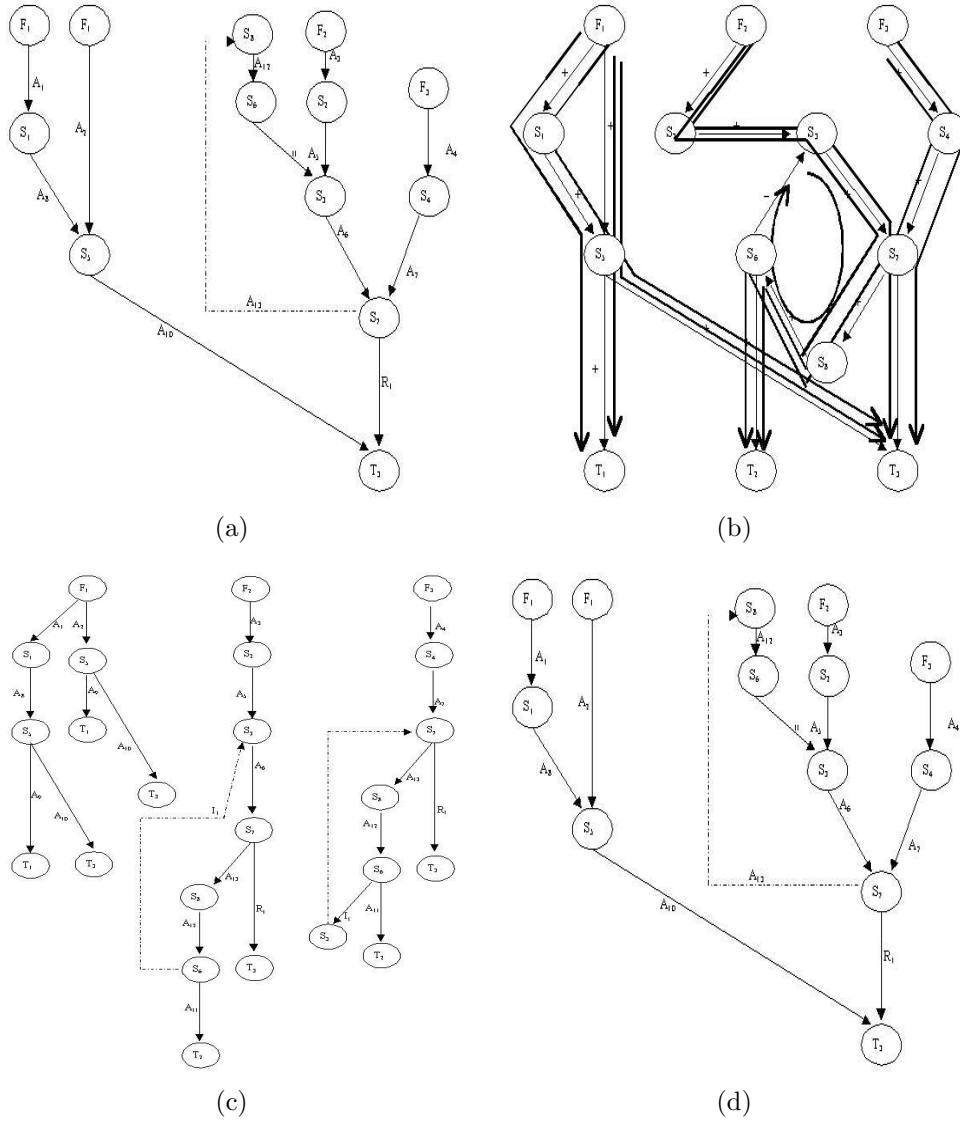
Figure 5.1:   (a) Graph representation of a signal map, containing only simple
effects such as A activates B or A inhibits B and monomolecular reactions. The
substances are already classified in factors: F_1-F_3, intermediaries: S_1-S_7 and tar-
gets: T_1-T_7. The activation are depicted with arrows A_1-A_{13}. The inhibition is
represented with an arrow I_1. The reaction is represented with an arrow R_1. (b)
Routes and a cycle in the previous graph. The activation are labelled with a plus
sign, the inhibition – with a minus sign and the reaction has no marker. (c) Tree
representations of the subgraphs of the original network. Each factor is root in a
such a tree. Going from the root to each leaf (which is a target), the signalling
routes are reconstructed. We learn which targets are affected by a given factor.
(d) Tree representations of the subgraphs of the original network. Target T_2 is
root in a such a tree. Going from the root to each leaf (which is a factor), the
signalling routes are reconstructed. We learn which factors are affecting a given
target.

without cycles, but detecting it will be helpful for a better understanding of the system. In Fig. 5.1(d), one can see on which routes target T_3 is affected by factors F_1, F_2 and F_3, and the cycle existing on the route from factor F_2 is also pointed out.

The most suggestive description of the network, in the present context, is provided by the balanced routes, the incomplete routes and the cycles. The cycles enrich the view. They give us information about feedback phenomena, which are wide-spread in biology.

In Appendix G, we give the procedure for extracting from the given graph the tree having as root a given node. Several details were omitted for brevity's sake. It is based on a breath first visiting method [Tom97, Knu74]. The tree structure is depicted in figure 5.2(b) and its base unit is the tree *node* (see figure 5.2(a)). For each *node*, only one of its *sons* and one of its *brothers* are saved in memory [CLR90, AJ83]. Note that each *brother* of a *son* of a *node* is, in fact, also a *son* of that *node*. Such a structure allows us to store the whole tree. Thus, we do not need to create a list of sons for each node. The aim is to minimize as much as possible the memory requirement, without increasing the information processing time. Obviously, the *root* can only have a *son* rather than a *brother*. The leaves have no *son*, while they may have *brother*s. An additional data structure is that of *father* which helps when we search for cycles.

The names used in the procedure are self-explaining. *UnexploredNodesList* is the list of the node pairs not yet analysed (*node, father of this node*). *CreateListUnexploredNodes(CurrentNode)* is a function that has as arguments the *current node* and the *direction*. We set the *direction* 1 if the *root* is a factor and the tree is constructed towards the targets, and –1, if the *root* is a target and the tree is assembled towards the factors. *CreateListUnexploredNodes* adds, to the existing *UnexploredNodesList*, the arrow starting from the *current node*, if *direction* is 1 and the arrow ending in the current node, if *direction* is -1. As long as this list is not empty, the function *ExtractNode(UnexploredNodesList)* extracts a *node*. The function *CheckCycle(CurrentNode, Node, Tree, direction)* verifies whether the current arrow (stored as its extremities: *CurrentNode, Node*) closes a cycle with the arrows which are already part of the tree. If the answer is yes, the function *NewCycle(CurrentNode, Node, CycleList)* checks whether a new cycle was detected. The new cycle is added to the *Cycles list*. If the arrow does not closes any cycle, the function *OldArrow(CurrentNode, Node, Tree, direction)* checks whether this arrow has already been added to the tree. If not, it is added as *son*, if the *current node* has no son, or as *brother*, if a *son* already exists. A new *node* is extracted from the *UnexploredNodesList*.

Figure 5.2: **(a)** The basis structure in trees: node. For each node, there are important the following fields: father, son, brother. **(b)** Method for storing a tree, based on the above presented fields. The root has no father and no brother. The leaves have no son. If a field is empty, it is presented in a grey-shaded box.

5.1.2 More complex networks

Multipart activations such as "A and B need to be present simultaneously to activate C" are also frequent in signalling maps. For example, to activate the IP_3 receptor in the membrane of the endoplasmic reticulum, both IP_3 and calcium ions are necessary (cf. [Gol96]). Due to these several input substances, a simple graph-based model is not any more suitable. Two types of nodes are necessary, one to model the metabolites and another – the activations and/or reactions figure 5.3. A convenient method for representing these features is provided by the formalism of Petri nets [Rei85]. A single-headed arrow connects two nodes of different types: places (represented by ellipses) and transitions (represented by rectangles). Commonly, also biochemical reactions converting substrates into products are involved in these processes. Due to the presence of multi-molecular reactions and multi-part effects, frequently, several arrows start in different places, while ending in the same rectangle and others start in the same rectangle, while ending in different places.
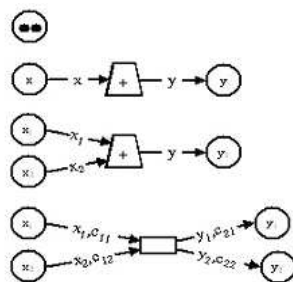
Figure 5.3: Petri net components necessary to describe signalling maps, containing only activation and reactions: **(a)** places stand for substances; tokens represent their number of molecules. They could be black (inactive form) or grey (active form). Transitions represent: **(b)** simple activations **(c)**multipart activations **(d)** biochemical reactions.simple activations

Due to the new, more complex form of a route, which is not anymore linear, to estimate the total effect becomes a very challenging problem. For example, if A is activated by an initial factor F1 and B is inhibited by an initial factor F2, we can determine the total effect on C (activated by A and B, see above) only if the extent of activation and inhibition is known quantitatively. Thus, a qualitative, binary decision (activation or inhibition) is no longer possible. Therefore, we restrict our analysis here by treating only activations and reactions. As mentioned above, the difference between activation and reaction is that the activator is supposed to remain in the system also

after activation. Indeed, in reality, the activator binds to the inactive form of the substrate, but it is assumed that it is released sufficiently fast so that it can act further in other effects or reactions. As in the case of monomolecular effects and reactions, a label helps us to take care of how each kind of arrow works in our model. In addition, the arrows are also labelled with positive integers, because usually the reactions have stoichiometric coefficients. These numbers are, in the Petri net literature, called weights and state how many tokens of each substance sort are "swallowed" by a given transition, and respectively, how many are produced. Petri nets have also been used in the modelling of metabolic networks [Hof94, RLM96, OBJO&D01]. Figure 5.3 illustrates the correspondence between the usual components of signalling maps and a Petri net apparatus.

For each substance, there are three states: first - when the substance is absent, second – when it is present, but in inactive form, third – when the substance is present in active form. The first state is modelled by a place without tokens, the second and the third correspond to a place with black tokens and grey tokens, respectively.

The places stand for factors, intermediaries and targets. At this moment, we assume that the factors are only substances, rather than medium conditions. As mentioned above, all kinds of effects and the reactions are modelled using transitions. In the Petri net literature, there is a somewhat imprecise terminology in that the term "transition"refers to the nodes represented with rectangles. However, a transition in the sense of transformation between places includes, of course, also the incoming and outcoming arrows. The place where the incoming arrow starts is called input place of the transition under consideration and that where the outcoming arrow ends is called output place.

Consequently, each output place of an effect must have at least so many black tokens as the label of the arrow pointing to it indicates, to allow the transition to fire. These tokens could be provided by a reaction, if such a reaction exists, or have to be part of the initial configuration. Importantly, the reaction can provide both already active products (grey tokens) and inactive products (black tokens). In the input file, one can mark the active form with a prime (') after the substrate name. Therefore, we shall have cases when a reaction producing an inactive substrate S and an activation that activates it are both required, but also cases where the reaction producing the active substrate S is sufficient.

To make a distinction between reactions and activations in the graphical description, we use rectangles only for the former and trapeziums – for activations. Arrow labels consist in pairs of a number (the stoichiometric coefficient) and a colour (grey or black). The second term of the pair specifies

the colour of the tokens that are required to be in the input places and of those are consumed or produced by a transition. For activations, the colour label of both incoming and outgoing arrows must always be grey because the active substance results in the activation of an initially inactive substance. Therefore, the colour label can, in the activation case, be omitted. In both, activation or reaction, for each place involved in a transition, a number of tokens equal to the number which labels the arrow starting from the considered place and ending in the given transition has to be present in the corresponding place to allow the transition to fire. Due to the distinction made between reactions and effects, during the algorithm, the tokens will have a different outcome. When the transition stands for a reaction, for each input place, a number of tokens equal to the first label number and having the colour indicated in the label will be deleted from the tokens set at that place. If the transition is an effect, neither the tokens number nor the colour will be changed. As for reactions, the number of tokens having the colour given in the second label member for each output place will increase by the label number. In the activation case a number of existing black tokens equal to the label number will become grey.

In fact, the transduction of a signal is composed of the firing of its transitions a specified number of times. This leads to a "migration" of tokens through the network. For sending the signal from factors to targets, a certain number of tokens has to be in each place. The vector containing the token numbers for each place between two transition firings represents a configuration (marking) of the system. The configuration before the first transition firing is the initial configuration. The minimal configuration that allows the firing of all transitions of a given route is a sufficient initial configuration. The configuration after the last transition firing of a given route is the final configuration. We do not fill places with tokens at the very beginning, because no prior information indicates how many tokens are necessary to go through each possible route. The transitions are added one by one to each route and, during these operations the number of tokens that has to be present at the beginning as well as the final configuration are calculated *a posteriori*.

The Petri net based description of complex signalling networks is a generalization of the classical graph-based description of simple networks. When we want to find all the targets that can possibly be reached starting from a given initial factor, the structure that occurs is not anymore a simple tree. To be able to disentangle the web including the three layers made up of initial factors, targets and intermediaries, the presence of cycles or the absence of reactions, it is essential to extract subnetworks with certain properties. We introduce the following concepts. **A place** is said to be

structurally balanced if the number of incoming grey tokens equals the number of outgoing tokens. A **place N** is **totally involved in a route** between one initial factor and one target if either there is a succession of transitions $t_1, \ldots \; t_m$ (called **transitions totally involved in a route**), such that t_1 has the initial factor as input place, t_m has the target as output place and there is an integer i, $1 < i < m$ such that t_{i-1} has N as output place and $_i$ has N as input place, or N is the initial factor or the target. Let us exclude reaction R2 from the network depicted in figure 5.4(a). In the new network, N', the nodes 1, 2, 3 and 5 are totally involved in a route between the initial factor 1 and target 5. The nodes 0, 2, 3, 4, 6 and 8 are totally involved in a route between the initial factor 0 and target 8.

Let TN_1 and TN_2 be the sets of places totally involved in the route from initial factor F to targets T_1 and T_2.

Targets T_1 and T_2 have an interrelation of degree 1 if, for the same initial factor F, there are places $M_1 \subseteq TN_1 – TN_2$ and $M_2 \subseteq TN_2 – TN_1$ being final places of one and the same transition. That is, this transition connects to one output place that is only involved in a route to T_1 (rather than T_2) and to another place only involved in a route to T_2. This means that there is a branching in a such a way that a signal running from F to T_1 may also arrive at T_2 and vice versa.

Targets $T_1 \ldots T_m$ have an interrelation of degree 1 if there is a permutation σ of $\{1 \ldots m\}$ and an integer $i, 1 \leq i \leq m$ such that each of the sets $\{T_1 \ldots T_{\sigma_i}\}$ and $\{T_{\sigma_{i+1}} \ldots T_{\sigma_m}\}$ have degree 1 of interrelation and denoting $TN_{1-\sigma_i}$ and $TN_{\sigma_{i+1}-\sigma m}$ the set of places totally involved in the route from initial factor F to targets $T_1 \ldots T_{\sigma_i}$ and $T_{\sigma_{i+1}} \ldots T_{\sigma_m}$ respectively, there are $M_{1m} \subseteq TN_{1-\sigma_i} - TN_{\sigma_{i+1}-\sigma m}$ and $M_{m1} \subseteq TN_{\sigma_{i+1}-\sigma m} - TN_{1-\sigma_i}$ being output places of one and the same transition between F and $T_1 \ldots T_{\sigma_i}$ and $T_{\sigma_{i+1}} \ldots T_{\sigma_m}$. The empty set and the set containing a single element are considered to have an interrelation of degree 1. In the network N' extracted from figure 5.4, the nodes 5, 8 and 15 have an interrelation of degree 1, because the nodes 8 and 9 are output places of transition R_3 and the nodes 4 and 5 are output places of transition R_0.

A **simple route** with respect to a set Θ of targets with interrelation of degree 1 and its corresponding minimal set of factors Φ must contain:

1. the targets from Θ and the factors from Φ,

2. the set of transitions and places totally involved in routes between pairs of one factor from Φ and one target from Θ, so that each place, except the factors and targets themselves, is structurally balanced,
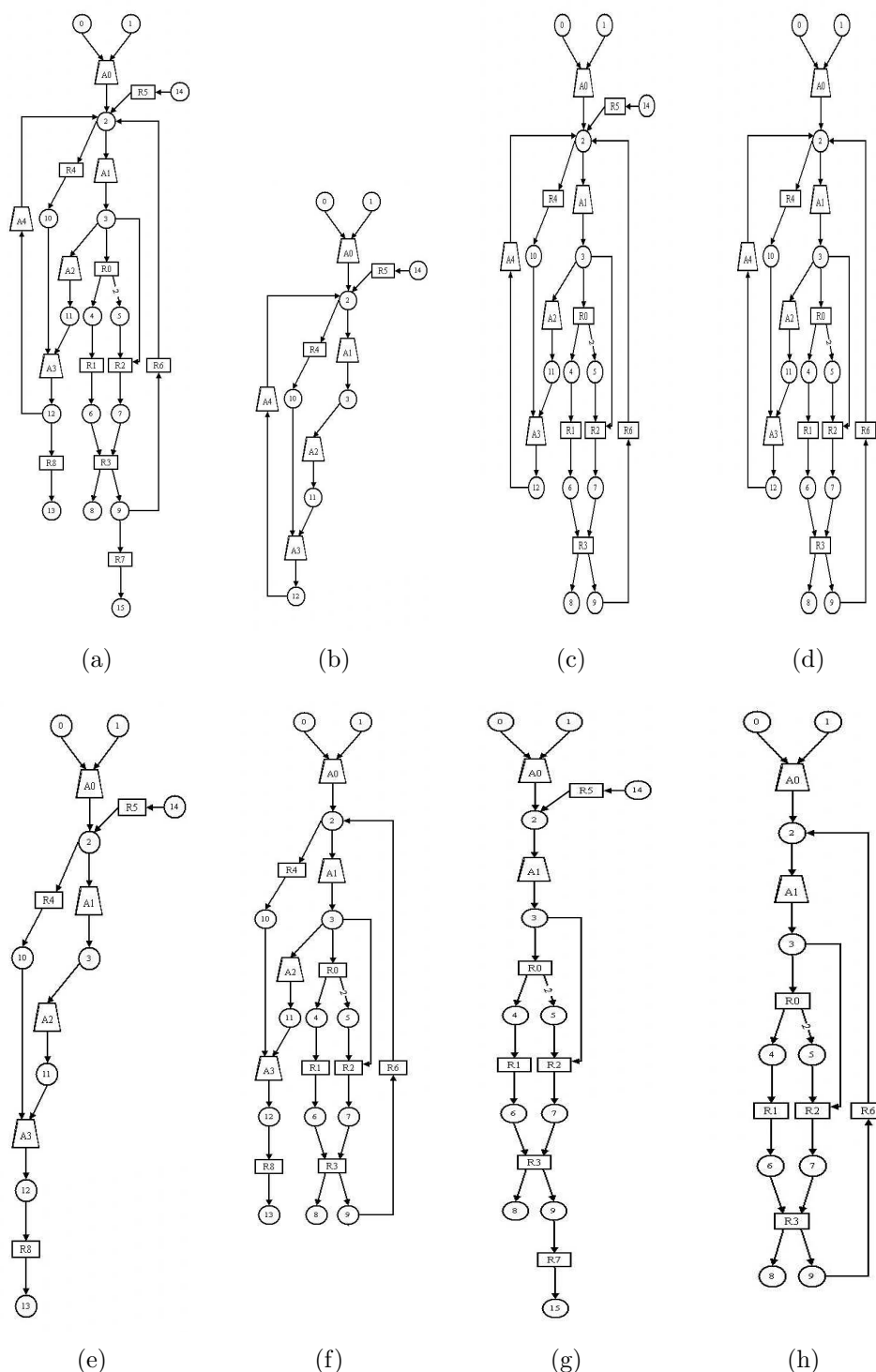
Figure 5.4: **(a)**Theoretical example with white circles as substances (factors – 0, 1, 14, intermediaries – 2, 3. 4, 5, 6, 7, 9, 10, 11, 12, targets – 8, 13, 15), one-line-framed rectangles as reactions R0- R7 and two-lines-framed rectangles as activations A0 –A4. The arrows are labelled only if the label is not 1. Here we detect four routes. Factors 0 and 1 act always together. Factor 14 often joins them. By detecting the conflict routes, the supposition can be derived that a reaction producing intermediaries 4 or 6 is missing, or a reaction consuming intermediaries 5 or 7 is missing. Each target is hit separately, exception targets 8 and 15 are hit simultaneously on the incomplete route. **(b)** balanced cyclic route: {2 R5, R4, A4, A3, A2, A1, A0}, **(c)** conflict cyclic routes: {R5, R2, R0, R1, R3, R6, R4, A4, A3, A2, 2 A1, A0}, **(d)** conflict cyclic routes: {2 R2, 2 R0, 2 R1, 2 R3, 2 R6, R4, A4, A3, A2, 3 A1, A0}, **(e)** balanced route: {R5, R4, R8, A3, A2, A1, A0}, **(f)** conflict cyclic routes: {R2, R0, R1, R3, R6, R4, R8, A3, A2, 2 A1, A0}, **(g)** conflict route: {R5, R2, R7, R3, R1, R0, A1, A0}, **(h)** conflict cyclic routes: {R2, R6, R3, R1, R0, A1, A0}

3. input or output places of a totally involved transition, even if they are not totally involved places.

A simple route may also contain:

1. transitions that have, as input places, places also belonging to the route, so that these places become structurally balanced, and the output places of such transitions,

2. transitions that have, as output places, places already belonging to the route, so that these places become structurally balanced, and the input places of such transitions.

The minimality condition has to be understood in the following sense: If a place totally involved in a route is deleted, no signal can be transmitted anymore. Considering again the network N', simple routes between the initial factors 0 and 1 and target 5 are the route $\{A_0, A_1, R_0\}$, the route $\{A_0, A_1, R_0, R_1\}$, the route $\{A_0, A_1, R_0, R_1, R_3\}$ and also the route $\{A_0, A_1, R_0, R_1, R_3, R_7\}$.

A **simplest route** with respect to a set of targets with interrelation of degree 1 and its corresponding minimal set of factors Φ contains at least the targets of Θ and at least the factors of Φ and all places and transitions belonging to this route are totally involved in this route and are, except the factors and targets themselves, structurally balanced. So a simplest route between the initial factors 0 and 1 and target 5 is only the route $\{A_0, A_1, R_0\}$.

In the case of multipart reactions/activations considered in this section, we should refine the concept of **balanced route**. We do so by invoking that it has a set of targets with an interrelation of degree 1 as final points, a minimal set of initial factors , and a set of intermediaries which are all structurally balanced. The route $\{A_0, A_1, R_0, R_1, R_3, R_7\}$ is an example of a balanced route between the initial factors 0 and 1 and target 5, but also between initial factors 0 and 1 and target 5, 8 and 15.

Place N_1 is said to be **situated before place** N_2 $(N_1 < N_2)$ if there is a succession of transitions $t_1, \ldots t_m$, such that t_1 has N_1 as input place, t_m has N_2 as output place and the intermediaries are balanced. In our example, it is obviously that node 0 is before node 2, which is before node 3. **Place** N_1 is said to be **situated after place** N_2 $(N_1 > N_2)$ if there is a succession of transitions $t_1, \ldots t_m$, such that t_1 has as input place N_2, t_m has as output place N_1 and the intermediaries are balanced. We can also say that node 3 is after node 2, which is after node 0. A **cycle** is a succession of transitions whose set of places has a subset Q with the property that for each pair $(N_1,$

N_2) of places from Q, there is a succession of transitions such that $N_1 < N_2$ and another succession such that $N_1 > N_2$. Therefore, A_1, A_2, A_3 and A_4 form a cycle, as well as R_4, A_3 and A_4. Cyclic routes with the targets set being empty affect the routes that contain a part of the cycle. The balanced routes and the balanced cyclic routes are the maximal simplest routes in the network with respect to the balancing property. A **balanced cyclic route** is a route having a set of targets with an interrelation of degree 1 as final points (that may be empty), a minimal set of initial factors, and all intermediaries are balanced, with the exception of one or several places where the cycle closes. In Fig. 5.4(b), the route $\{A_0, A_1, R_4, A_2, A_3, A_4, 2R_5\}$ is a balanced cyclic route. If a cycle contains only activations but no factors, it cannot normally transmit a signal due to the lack of input signal. However, strong fluctuations in the system may cause a metabolite to become active. In this case, all metabolites involved in the cycle become and keep being active, which serves to the amplification of the signal. Thus, it is of interest to determine such cycles.

Let us consider a partial order relation on the set of subnetworks that we have defined. We say that **network $\Xi_1 < \Xi_2$** if the set of transitions of Ξ_1 are included in the set of transitions of Ξ_2. In accordance with this definition, the route $\{A_0, A_1, R_0\}$ is smaller than route $\{A_0, A_1, R_0, R_1\}$. Then, all balanced routes, balanced cyclic routes represent a sup Σ, where $\Sigma=\{$ordered simplest routes from F to subsets of targets $\{T_1 \ldots T_m\}$ with an interrelation of degree 1$\}$. Such a balanced route α from F to $\{T_1 \ldots T_m\}$ is a simplest route because all the intermediaries are totally involved in the route and all of them are balanced, so it is part of Σ. Moreover α is greater than all other routes from Σ, because each other route β hits a subset of $\{T_1 \ldots T_m\}$, so they must have at least one arrow less. Let us consider the ordered simplest routes from initial factors 0 and 1 to the subsets of targets 5, 8 and 15 which have an interrelation of degree 1. These are: route $\{A_0, A_1, R_0,\}$ route $\{A_0, A_1, R_0, R_1, R_3\}$ and also route $\{A_0, A_1, R_0, R_1, R_3, R_7\}$. Their supremum is $\{A_0, A_1, R_0, R_1, R_3, R_7\}$ and it coincides with the balanced route mentioned above.

Detecting the cycles can give us valuable information. In signalling networks, a very frequent process is that of feedback regulation. This mechanism has a vital importance in any living organism. If some substances are produced in a larger amount than required, a negative signal is sent back to reduce the unnecessary or even dangerous excess production. This is a negative feedback. The cycles that occur in signalling networks mainly have such a function. If we want to enhance the synthesis of a desired product, it is not enough to know which factors have to be varied, but also if some feedback loops prevent us to do it. Using genetic engineering, one can remove the

feedback loop to reach the aim (cf. [CBHC95]). On the other hand, positive cycles are in accordance with amplification phenomena and can lead to oscillations. For example, a prominent mechanism leading to calcium oscillations is the positive feedback from cytosolic calcium on its release from the endoplasmic reticulum (cf. [Gol96]).

Now we consider the special case that all arrows of a subnetwork under consideration represent reactions. A **place** is **flux-balanced** if it is consumed in the same amount in which it is produced. An **accumulating place** is consumed in a smaller amount than it is produced. A **deficit place** is consumed in a greater amount than it is produced. Depletion and accumulation are allowed only in conflict states. A **conflict route** occurs when an intermediary accumulates or depletes due to a missing reaction. It is impossible to equilibrate such an intermediary and the accumulation or depletion does not occur due to a cycle containing the accumulating place. A **cyclic conflict route** contains at least one unbalanced intermediary due to a missing reaction and one due to the transition closing the cycle. All other intermediaries are balanced. If in our (cyclic) route, there is at least one intermediary that accumulates or depletes, it might be that some reactions are missing in the system. Detecting the possible accumulation and depletion points in the system, one can draw the experimentalists' attention to the search for hitherto undetected reactions.

However, not always when gaps exist in signalling maps, (cyclic) conflict routes occur. The network topology plays a decisive role. It is essential that a particular state arises, which we describe by "conflict state". To clarify it, let us assume that a subsystem in our map looks like in figure 5.5. Note that substance P_2 is produced with the stoichiometric coefficient of 2. Both reactions R_2 and R_3 could have some coproducts. Their stoichiometry is not interesting for us. What is important is that P_3, P_4 and maybe other metabolites are substrates in reaction R_4, each with coefficients 1. Obviously, after firing R_1, we will have one molecule of P_1 and two molecules of P_2. This configuration allows reaction R_2 to fire once and reaction R_3 – twice. We will obtain one molecule of P_3 and two molecules of P_4. Now, we have enough molecules of P_3 and P_4 to fire R_4 once. But so, one molecule P_4 remains in the system. R_1 and R_2 could fire again to produce the second molecule of P_3. These will permit R_4 to fire the second time, but another molecule P_2 will occur. The longer the process runs, the more molecules of P_2 or P_4 will accumulate in the system. So P_2 or P_4 is an accumulation place. This is what we call a conflict case. One may argue that signalling is a time-dependent process and need not reach a steady state. Thus, if some deposit is built up (of P_2 or P_4 in our example), it could be consumed later. However, due to the unbalanced stoichiometry, this would lead to depletion or accumulation
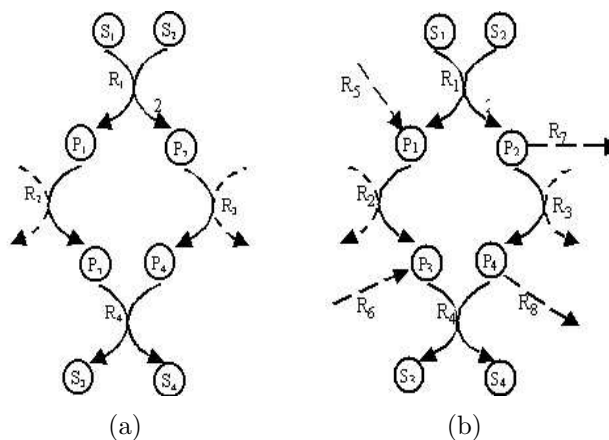
Figure 5.5: Simple example illustrating conflict cases. Dashed arrows indicate possibly occurring side reactions. **(a)** Conflict cases: In reaction R_1, one molecule S_1 and one molecule S_2 are consumed and one molecule P_1 and 2 molecules P_2 are produced. R_2 consumes one P_1 and produces one P_3. R_3 consumes one P_2 and produces one P_4. R_4 consumes one P_3 and one P_4 and produces one S_3 and one S_4. This will lead to accumulation of P_2 or P_4. **(b)** If R_5 produces one P_1 or if R_6 produces one P_3, R_4 can fire twice, consuming two P_3 and two P_4. The conflict is resolved. If R_7 consumes one P_2 or if R_8 consumes one P_4, R_4 can fire only once, but no substrate is remaining in the system. This is another solution of the conflict.

at other places. Thus, the only interpretation is that some reactions, which have to consume P_2 or P_4, are missing. To identify the conflict cases is also important for the algorithm itself, because upon trying to equilibrate the flux or transfer of information, infinite loops will occur.

The conflict cases can be resolved by extra reactions, as is illustrated in figures 5.5(a) and 5.5(b). If there are some other fluxes coming in the considered reactions set or going out, the required quantity of P_1 or P_3 can be produced by R_5, or, respectively R_6 or the surplus of P_2 or P_4 can be better used in R_7, respectively R_8. Imagine now that reaction R_1 produces one molecule of P_1 and one molecule of P_2. Even if, in reality, a reaction that has to produce one P_1 or P_3, or it has to consume P_2 or P_4 is missing, due to the network's topology, this does not degenerate into such a conflict state, so the gap will not be observed.

The conflict routes and cyclic conflict routes represent a sup Σ, where $\Sigma=\{$ordered simplest conflict routes from F to subsets of targets $\{T_1 \ldots T_m\}$ with an interrelation of degree 1$\}$. Such a balanced route α from F to $\{T_1 \ldots T_m\}$ is a simplest conflict route because all the intermediaries are totally involved in the route and all of them are balanced, so it is part of $T\Sigma$.

Moreover α is greater than all other routes from Σ, because each other route β hits a subset of $\{T_1 \ldots T_m\}$, so they must have at least one arrow less.

5.2   Algorithm and implementation

Our algorithm is based on the backtracking strategy combined with some appropriate heuristics. In accordance with the general backtracking procedure, the solution represents an array $(X_1, X_2, \ldots, X_n, \ldots)$. Its dimension is not *a priori* fixed. X_i stands for the transition added to the route at step i of the algorithm. To avoid running time problems, the domains D_i, in which each X_i takes values is considerably restricted [LG86]. Immediately after the reading of the input data, we determine the sets of neighbour transitions D_p for each place p in both directions. Another computational artifice is to consider at step i a pair of a place and a transition. Actually, our solution looks like $((p_1, X_1), (p_2, X_2), \ldots, (p_n, X_n),\ldots)$, where p_i is the place visited at step i and X_i, takes values from D_{pi}. Thus, the search space is very much diminished. When such a pair is added to the solution array, some conditions have to be fulfilled. The tokens have to be sufficient in number for each place. If not, the initial sufficient configuration is adapted, if it is not a conflict case. If it is, one must check whether there are additional reactions which can resolve the conflict. If so, the algorithm continues, but if not, the program is prepared to find and store a conflict route. If the transition closes a cycle, the program is prepared to store the solution in the corresponding stack.

Let us assume that we want to find the targets that are effected by factor F_1. We shall also discover which other factors work together with F_1. Let us take a look on the basic procedure. At the first step, our partial solution is $((F_1, X_1))$. X_1 is an element of D_{oF1}, which is the set of transitions going out from F_1. We set the tokens configuration, we put the input places of X_1 in a waiting stack, which has to be expanded in the direction "*up*". This means that their domains will be "*in-domains*". Also the output places of X_1 are put in a waiting stack, which has to be expanded in direction "*down*"; their domains will be "*out-domains*". While not all the routes were found and while the current route is not yet a solution, the current configuration is updated, a place that has to be expanded is chosen, a transition from its domain is selected. If the total solution is found, it is stored and the transition added last is taken out from the current (partial) route. Another transition from the current place domain is added. This procedure is repeated until all the solution routes are generated.

The algorithm has the following advantages. Because variables domains
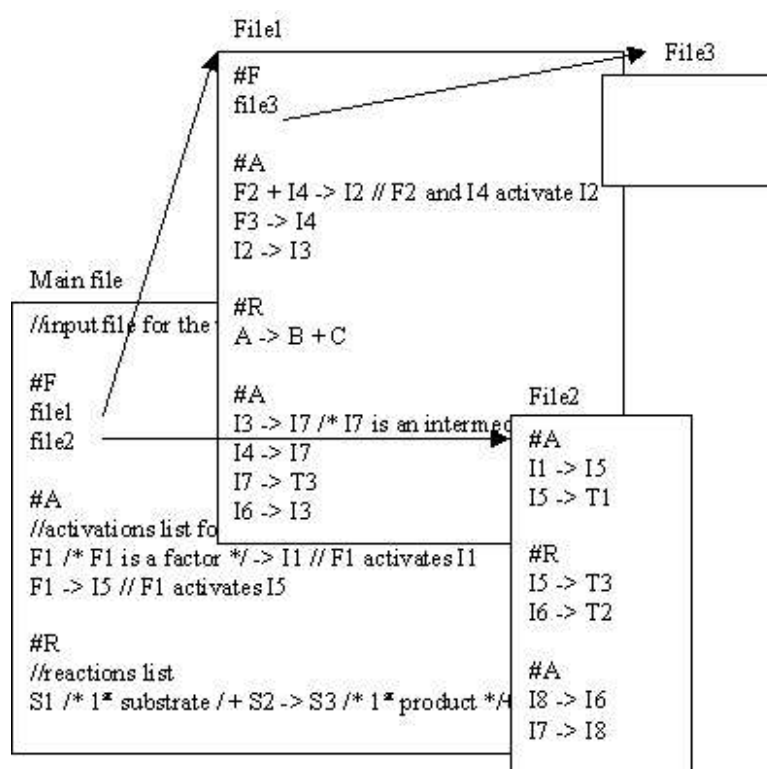
Figure 5.6: Input file caracteristics: Comments in C++ manner: starting with //
and ending at the end of current line starting with /* and ending with */ comments
insertion in equations. Repeatedly clauses marked with #A (for activations) and
#R (for reactions). Data distributions in several files, which can be appeled in
clause #F. Equations writing on several lines.

are neighbour effects and reactions, they do not have large dimensions. The normal large complexity of a backtracking technique is considerably reduced. The routes are generated in a random order. There is no failure case, each choice of a transition leading to a solution. Thus, each examined route brings an evaluation. It gives a balanced route, it identifies a cycle, or it finds possible gaps.

We have implemented the above presented algorithm in the SigNetRouter application, using C++ programming language. The required input data are taken from text files. In the following, we shall present the current input file features (Fig. 5.6). For each transition type, there are sections marked by **#R** (for reactions) and **#A** (for activations). The existing system can be extended by including additional sections of this type or, even data files. Their name have to be written in a **#F** clause. Comments in the style of C++ programming are allowed. They facilitate the later understanding of an existing input file, in order to reuse it. One input file is already declared by default. It could be replaced with another file, specified by the user and possibly having another location. Certainly, if syntactic mistakes occur in the input file, an error and its description will be reported. The user will be guided as close as possible to the line where the ambiguity or mistake seem(s) to be and is invited to review it.

Once the data were read, some preliminaries are accomplished. An adjacency matrix is built to store the neighbourhood relation between metabolite and transitions. A precedence matrix is calculated to identify the network's cycles. A conflict matrix is constructed to detect conflict cases and to prevent infinite loops, which could occur if the non-self-unlocking conflict places are further explored.

The precedence matrix is, in fact, a half matrix $(P_{ij})_{i=\overline{1,n}, j=\overline{i,n}}$, where n is the number of places. The values above the main diagonal are enough to detect the cycles. Its elements are quadruples $P_{ij} = (relation\_edge, intermediar\_type, edge, type)$. During the algorithm the meaning of this components changes. This allow us to spare other variables. At the first step, *relation_edge* takes a certain negative value (we arbitrarily set it -12), if it is not known whether i is before or after j on a route. If the sequence on a route is (i, j), it takes another certain value (-11), and in the opposite case, -1. *Intermediar_type* is –10, if no route between i and j is known or if between them there is an arrow in our network, no matter in which direction.

The precedence matrix is constructed step by step. If at the step s, we know that i is before j and j is after k, at the next step, we know whether i is before k. If so, k will be stored in the *Intermediar_type* field (otherwise i). *Edge* takes the value –20, if no arrow binds i and j. Otherwise, it stores the arrow identifier. Type is –20 if there is no arrow between i and j, t6 if the

arrow stands for an activation and –5, for a reaction. When this quadruple already stores information regarding a certain direction and also information concerning the opposite reaction is available, the first two fields correspond to the forward reaction or activation, whereas the last two fields, to the backward step.

Let us consider the examples represented in 5.7: In the first example, S_1 is before S_2 on a route composed by activation A_1, but is after S_2 on a route composed by R_1. Therefore, $P_{S_1 S_2} = (R_1, -5, A_1, -6)$. We recall that the subscripts of R_1 and A_1 are for the algorithm simply numbers and they carry no information about their types. In the second example, S_1 is before S_2 on a route composed by activation A_1, but is after S_2 on a route going through S_3. Therefore, $P_{S_1 S_2} = (-1, S_3, A_1, -6)$. As it is shown in the third example, if S_1 is before S_2 on a route traversing S_4, but is after S_2 on a route going through S_3, $P_{S_1 S_2} = (-1, S_3, -11, S_4)$.
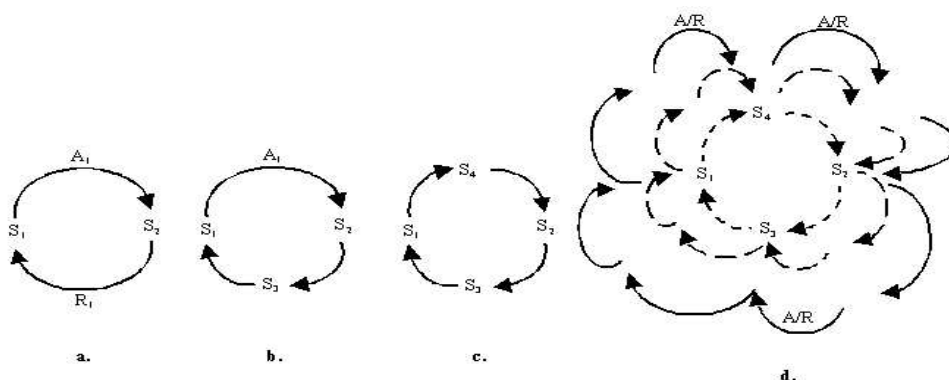

Figure 5.7: **(a)** S_1 is before S_2 on a route composed by activation A_1, but is after S_2 on a route composed by R_1. Therefore, $P_{S_1 S_2} = (R_1, -5, A_1, -6)$. **(b)** S_1 is before S_2 on a route composed by activation A_1, but is after S_2 on a route going through S_3. Therefore, $P_{S_1 S_2} = (-1, S_3, A_1, -6)$. **(c)** If S_1 is before S_2 on a route traversing S_4, but is after S_2 on a route going through S_3, $P_{S_1 S_2} = (-1, S_3, -11, S_4)$. **(d)** Process of retrieving cyclic routes. Starting from information in the precedence matrix, in between each pair of nodes, a node from the original network is inserted in an iterative manner. In the last step, the corresponding reactions or activations can be inserted.

So, having a half-matrix with these quadruples as elements, we fill it, at the beginning, with arrows in the sense of Petri nets. As long as possible, we construct new relations based on the transitivity property. If S_1 is before S_2 ($S_1 < S_2$) and S_2 is before S_3 ($S_2 < S_3$), then S_1 is before S_3 ($S_1 < S_3$). Every time, the intermediary place is stored, thereby the cyclic routes can be retrieved from the precedence matrix. This retrieval process is represented

in 5.7(d).

The conflict matrix $(C_{ij})_{i=\overline{1,m},j=\overline{1,m}}$ gives the ratio between each two places connected by a reaction chain in the network. For example, if reaction R_1 consumes 2 molecules of substance S_1 and produces 3 molecules of S_2, the ratio is $C_{S_1 S_2}=\frac{2}{3}$. Obviously, if $C_{S_i S_j}=\frac{s}{p}$ and $C_{S_j S_k}=\frac{r}{t}$, then $C_{S_i S_k}=\frac{s \cdot r}{p \cdot t}$. In the conflict cases, there are two routes from S_i to S_k, and the ratios are different. In Fig. 5.5, for example, $C_{S_1 P_2}=\frac{1}{2}$, $C_{P_2 S_4}=1$, $C_{P_4 S_3}=1$, so $C_{S_1 S_3}=\frac{1}{2}$, but $C_{S_1 P_1}=1$, $C_{P_1 P_3}=1$, $C_{P_3 S_3}=1$, so $C_{S_1 S_3}=1 \neq \frac{1}{2}$. When in the conflict matrix such different ratios are calculated, the places involved in this conflict are marked. During the backtracking procedure, when these places have to be explored, their outgoing and ingoing transitions are counted. If these places are accumulating places and there is only one outgoing transition, or if these places are deficit places and there is only one ingoing transition, an incomplete route has to be built and stored. Otherwise, the conflict is solved on its own, and the backtracking strategy is proceeding smoothly.

The backtracking procedure that builds and classifies route after route having as start point a given place is given in Appendix H. It is written in pseudo-code and some details, which would make it difficult to understand, are omitted. *Startplace* stands for the factor or the target where the route has to start or to end. Because the route can be constructed forwards or backwards, depending on the type of *StartPlace* (factor or target) and also on the network's topology, we need to give the *Direction* as parameter. Although *Step* counts the route length, its purpose is to increase and decrease in a such a way to allow all the routes to be identified and to be a flag for finishing the procedure. At the beginning, the lists where all kind of routes will be stored has to be initialised. The procedure *InitialiseLists*() does this. Because the algorithm builds a route and then, deletes its last element to build another route, the intermediary states have to be memorized. *StoreCurrentState*(*Step*) is responsible for this operation. The *CurrentPlace* at the current *Step* is also initialised with the *StartPlace*. *TypeT* is a variable that stores integer values, depending on the transition type (reaction: 1, activation: -1, inhibition, effect, not known). As a function of *Step*, *Direction* and *TypeT*, the procedure *SetSolutionSets*() sets the lists of transitions from which it will choose the next transition to be added to the route. We use two *SolutionSets*, the first one - for all kind of effects and the second – for reactions. As long as *Step* is not yet zero and at least one *SolutionSet* is not empty, if we are on the backward way, we have to restore the precedent state. The procedure *RestoreState*() is doing this. The current state has to be memorised by *StoreCurrentState*(*Step*). A new transition to be added to the partial route is chosen by the procedure *ChoseTransition*. Transition type is memorized. The route obtained until now is verified. If it is a com-

plete route, the function *Solution*() will return 1. The current route will be compared with the already stored balanced routes. If it is a new one, it will be added to the balanced routes list and the information regarding initial and final configuration will be also managed. The same treatment applies to the conflict routes and cyclic conflict routes, which will be stored in the corresponding lists. If the partial route is not yet complete, a new place to be explored is chosen and the corresponding *SolutionSets* are set.

An example is given in Fig. 5.4. There are three factors (0, 1, 14) and three targets (8, 13, 15). The other places are intermediaries (2−7, 9−12). All of them are depicted with white cycles. The rectangles stand for reactions (R0−R7) and trapeziums - for activations (A0−A4). The arrows are labelled only if the label is not 1. Let us chose as start place the factor 0. We detect seven routes: A balanced route: {R5, R4, R8, A3, A2, A1, A0}, a balanced cyclic route: {2 R5, R4, A4, A3, A2, A1, A0}, a conflict route: {R5, R2, R7, R3, R1, R0, A1, A0} and four conflict cyclic routes: {R5, R2, R0, R1, R3, R6, R4, A4, A3, A2, 2 A1, A0}, {2 R2, 2 R0, 2 R1, 2 R3, 2 R6, R4, A4, A3, A2, 3 A1, A0}, {R2, R0, R1, R3, R6, R4, R8, A3, A2, 2 A1, A0} and {R2, R6, R3, R1, R0, A1, A0}. Factors 0 and 1 always act together. Factor 14 often joins them. It is supposed to miss a reaction producing intermediaries 4 or 6, or another consuming intermediaries 5 or 7. Each target is hit separately, except targets 8 and 15, which are hit simultaneously on the conflict route.

In the output file, one can find a classification of metabolites in factors, targets, and intermediaries. It is very unlikely that a factor can work on its own. The factors that could support it are given in the output file. The same is valid for targets. Also other targets may be influenced when only one is directly envisaged at the beginning. They are also displayed. One can view the routes on which a factor affects some targets and which are the targets. Conversely, also the factors which influence a given target and their routes are found. Cycles that occur in between are detected. Moreover, the sufficient initial and resulting final configurations are given in the output file. An example is shown in 5.2.

Sub-stances	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Initial conf.	1g	1g	0	1b	0	0	0	0	0	0	0	1b	1b	0	2b	0
Final conf.	1g	1g	1g	1g	0	0	0	0	0	0	1g	1g	1g	0	0	0

Table 5.2: Sufficient initial and resulting final configurations for the balanced cyclic route shown in 5.4(b)."g"stands for grey tokens, while "b"– for black ones.

5.3 An example: The B-Cell antigen-receptor signalling network

The signalling network associated with the B cell antigen receptor has attracted enormous interest [Cam99, DeF95, DeF97], due to its importance in living organisms. The antigen receptor complex of B cells recognizes and discriminates between various structures, detecting the infectious agents. This triggers the immune response of the organism against viruses, bacteria, etc.

Approaches to design signalling networks can also be found in the internet

(http://www.cellsignal.com/reference/pathway/BCR.asp,
http://www.grt.kyushu-u.ac.jp/spad/menu.html).

The latter database also includes information about to the BCR pathway. Some signalling schemes therein are, however, oversimplifications. The difference between reactions and activations is often not made clear. To understand the signalling events better, it is usually necessary to look into the original literature [Cam99]. Once the antigen is bound to the receptor, it becomes possible for Lyn to phosphorylate the receptor (in the scheme in the database, it looks as if Lyn triggers the signal). Syk also binds to the receptor and becomes active. Lyn activates Btk, which participates besides Syk in activating of the complex formed from Sos, Vav, PLCγ and Grb2. This complex can then activate Rac_GTP, RhoAGTP, Cdc45GTP and GTP_Ras, as well as DAG and IP$_3$ derived from PIP. IP$_3$ activates Ca^{2+}, which activates RelA, cRel, NFaB and CaM. CaM activates NFAT. DAG activates PKC, which activates Rap. Each Rac_GTP, RhoAGTP, Cdc45GTP can activate PIP5, PAK, JNK or p38. GTP_Ras activates Raf1, which activates MEK, which activates ERK. GTP_Ras can be also activated by SHC, which is activated directly by Syk. We think that a more appropriate representation is given in Fig. 5.8, using the Petri net apparatus explained in the previous sections. We take into account that the antigen activates the receptor. This enables the Lyn binding and an active complex CR1 is formed, becoming available for binding Syk so to form CR2. Our representation prevent ambiguities coming from hidden intermediary steps. It is formal and rigorous.

Running SigNetRouter on the input file formalising this scheme, we refine the conclusions of [Cam99]. Where they identified the PLCγ pathway, SigNetRouter finds four distinct balanced routes carrying the signal from the receptor to Rap and each RelA, cRel, NFab and NFAT, respectively. Twelve balanced routes correspond to the so called Rho-family pathway. Importantly, there are three balanced routes leading from the receptor to each final target: PiP5, PAK, JNK and p38. From the receptor to ERK on the so
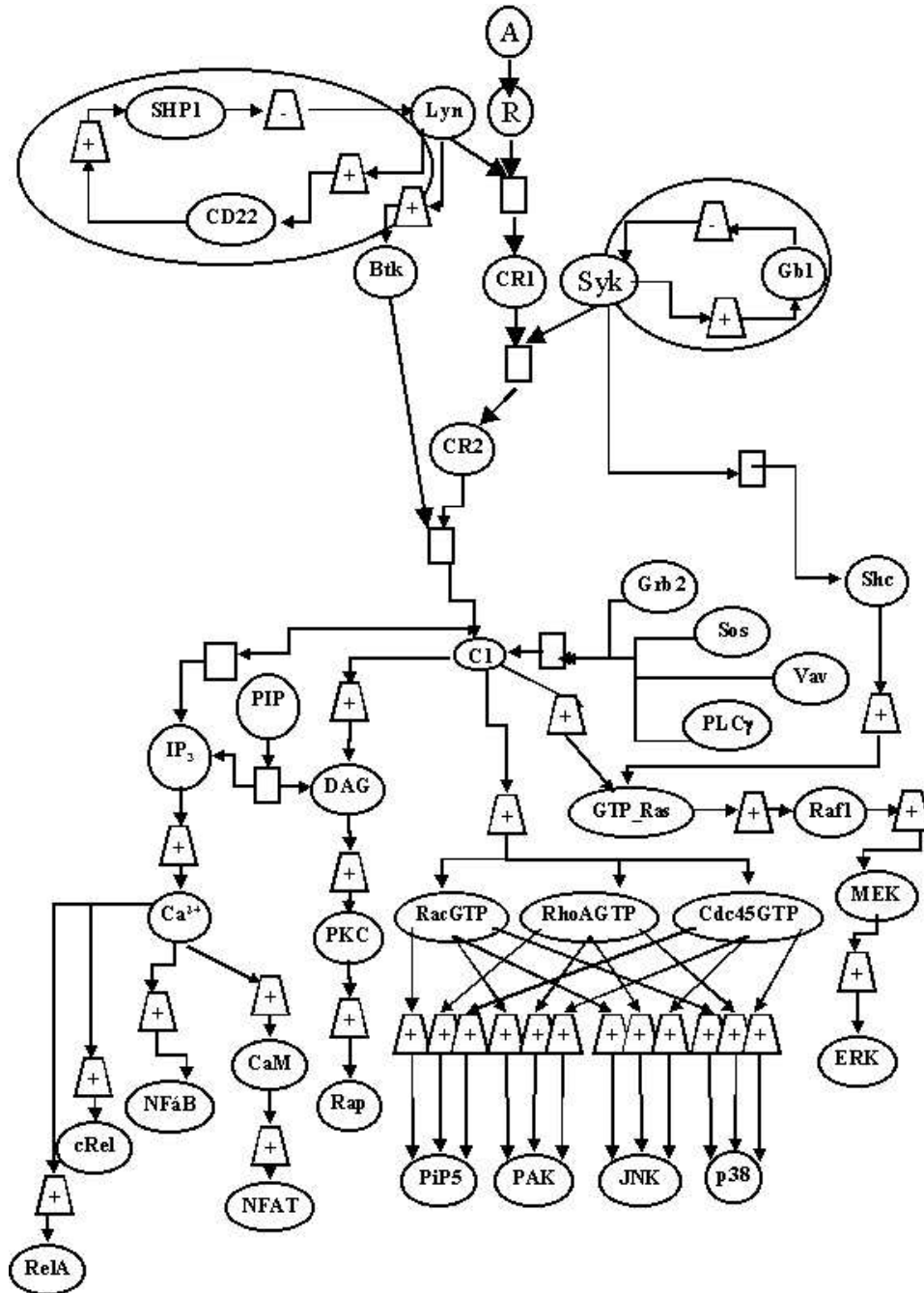
Figure 5.8: BCR Signalling network. A – antigen; R – receptor; Trapeziums – effects (+ activation; - inhibitions); Boxes – reactions

called Ras pathway, there are also two balanced routes. This shows us that the network is structurally robust to a certain extent because there are parallel routes (at least at its lower part). All these balanced routes that we found have a common part (from the antigen to the complex C1 - receptor activates Lyn, which activates Btk and Syk, which in turn activate complex C1). One could think that the network could be vulnerable to defective functioning of some intermediary steps between the antigen and complex C1. This is not, however, the case, because alternative routes exist between the receptor and the above mentioned targets. We did not bring them up, due to the presence of some inhibition in between – CD22 was shown to be a negative regulator of B-cell receptor signalling [NCO$^+$97] t which our model can not totally cover yet. It is obvious that their presence is very important and therefore, efforts have to be made to solve the problem of including inhibitions.

Chapter 6

Conclusions and discussion

6.1 The impact of multifunctional enzymes

Multifunctional enzymes (in the sense of enzymes with low substrate speci-
ficity) are ubiquitous in living cells. Prominent examples are alcohol de-
hydrogenase, aldolase, hexokinase, transketolase, uridine kinase, cytidylate
kinase, nucleoside diphosphokinase and adenylate kinase. Many other en-
zymes, which are commonly considered as monofunctional, actually catalyse
a number of side reactions, the so-called underground metabolism [DC98].
To our eyes, these facts have not always been duly taken into account in the
modelling of metabolic networks. In the first chapter (1) of the thesis, we
have investigated how to treat multifunctional enzymes in metabolic pathway
analysis.

We have shown that the various reactions that can be performed by a
multifunctional enzyme catalysing reversible bimolecular reactions are usu-
ally linearly dependent on each other. Usually, in the biochemical litera-
ture, only a set of linearly independent reactions is given. For example,
for transketolase, only **Tkt1:** R5P + X5P = S7P + GA3P, **Tkt2:** E4P
+ X5P = F6P + GA3P are used. However, the choice of an appropriate
set of linearly independent reactions of a multifunctional enzyme was until
now arbitrarily. For example, instead of the reaction **Tkt2,** the reaction
Tkt3: E4P + S7P = X5P + F6P could be taken with the same justifica-
tion as a "basic"reaction. The alternative chosen by a number of databases,
e.g. BRENDA (http://www.brenda.uni-koeln.de/) and partly in the litera-
ture is, especially for enzymes with a high number of functions, to indicate
the set of possible substrates and products without specifying the particular
reactions. For example, for transketolase, sometimes the set of donors of
the glycolaldehyde moiety and the set of acceptors is given (Bykova et al.,

2000). ExPASytENZYME (http://www.expasy.ch/enzyme/) indicates that many nucleoside diphosphates can act as acceptors in the nucleoside diphosphokinase reaction, and many ribo- and deoxyribonucleoside triphosphates can act as donors. Obviously, the number of pathways would be much greater than the number obtained in Section 4 if all functions of this enzyme were included.

When using overall reactions in pathway analysis, it is important to consider only linearly independent reactions. Otherwise, cyclic elementary modes are obtained which do not perform any net transformation and have to be cancelled afterwards. For example, consider the three reactions Tkt1, Tkt2 and Tkt3 in the system given in Fig. 1.1. If all the metabolites are internal, then we obtain one elementary flux mode: Tkt1, -Tkt2, -Tkt3, which does not perform any net transformation and has no biological meaning. Also, including NDK4 into the set of reactions for the system shown in Fig. 2.4, we will obtain a spurious elementary mode consisting only of enzyme NDK (with the functions NDK2, -NDK3, NDK4) and performing no net transformation at all. An algorithm for choosing, in a unique way, an appropriate set of linearly independent functions has been given. We suggest that such information could be included in metabolic databases.

In all examples we tested so far, the number of different functions of multifunctional enzymes as derived from the convex basis of the half-reactions system is equal to the number of linearly independent functions as given by the rank of the stoichiometry matrix. It will be a subject of future studies to prove that this is true in general for multifunctional enzymes involving at least one reversible reaction.

The main result given in the first chapter (1) is that the number of pathways (defined as elementary flux modes) does not depend on the level of description of multifunctional enzymes provided that the definition of elementary modes is applied correctly. This requires that enzymes are considered as the basic units in metabolism. We have to find a minimal number of enzymes rather than a minimal number of reaction steps. Note that the definition of elementary modes is based on the principle of genetic independence introduced by [SB88]. This says that upon decomposition of a mode into two other modes, no enzymes should be used in addition that are not used by the mode in question because the other modes then would be genetically independent. Formal application of the definition on the level of hemi-reactions would imply that the principle of genetic independence would hold for hemi-reactions, which is not the case. This point is also related to the concept of activity set introduced by [NnSVPI+97]. This is the set of all enzymes operative in a given state. Note that in the presence of multifunctional enzymes, this set involves fewer items than non-zero components are contained in the

vector $\mathbf{V}$ because several components refer to the same enzyme.

We have illustrated these theoretical considerations by several examples (a system made up of uridine kinase and other reactions of nucleotide metabolism, pentose phosphate pathway and interconversion of nucleoside triphosphates). Another example is the reaction system of monosaccharide metabolism studied in [NnSVPI+97]. In terms of overall reactions, this system gives rise to 296 elementary flux modes, while in terms of half-reactions, 866 modes would arise if the reaction steps were considered as basic units. Note that [NnSVPI+97] analysed the convex basis. This set involves eight basis vectors in both descriptions. Also for the other examples, we have found that the number of basis vectors is independent of the level of description even if the definition is applied to reaction steps. It can be shown that this holds for any reaction system because the basis vectors have the property that none of them is a non-negative linear combination of other basis vectors. The same property holds after translating the basis vectors of the half-reactions system into those of the overall reactions system. When a system only contains irreversible steps, the set of elementary modes coincides with the convex basis [SH94]. Therefore, such systems do not cause any problems with respect to a description on different levels of detail. As illustrated by an example in Section 2, the same is true as long as only monofunctional enzymes are reversible.

Both sets of fundamental pathways have their advantages: the convex basis usually involves a smaller number of pathways, but they are sometimes not uniquely determined [PSVNn+99] and they often do not cover all biochemically relevant routes ([SHWF02]. Moreover, they may change considerably upon addition or deletion of reactions. For example, if reaction $\mathbf{b}$ in Fig. 2.3(a) is deleted, the previous extreme vectors do no longer apply and the new extreme vector $\{\mathbf{a}, \mathbf{c}\}$ arises. The set of elementary modes is uniquely determined and involves the full set of potential basic routes in the system. Moreover, if enzymes are added to the system, new elementary modes may arise while the previously existing modes remain unchanged ([SHWF02]. Upon deletion of enzymes, some modes may disappear while none of the remaining modes will be changed. While the convex basis is always independent of the level of description, the elementary modes require a correct application of the definition in order that the same property holds for these.

Our analysis is based on several simplifying assumptions. For example, we have not taken into account that transketolase consists of two subunits and that each of them can catalyse, in addition to the two-substrate reactions mentioned above, also a one-substrate reaction in which a two-carbon unit is split from a donor molecule [FGS+01, BSM+00]. The two C2 moieties

are then combined to give erythrulose. Our analysis can be extended in a straightforward way to cope with dimeric enzymes.

In future studies, it is worth considering that different reactions of a multifunctional enzyme can proceed depending on conditions (e.g. availability of initial substrate). However, the general results of the present contribution will not be affected by this specification. Moreover, the approach presented here may be of interest for the modelling of enzyme evolution because it is generally assumed that the multitude of present-day monofunctional enzymes evolved from a smaller number of enzymes with broader substrate specificity.

6.2   Defects and "hijacking" of enzymes in nucleotide metabolism

As mentioned above, nucleotide metabolism is a very important part of each organism. It provides the building-block necessary in DNA and RNA synthesis. It supports the immune response. NAD^+, FAD^+, Coenzyme A, cAMP, cGMP, and other important substances need nucleotide metabolism also. Some organisms contain the whole system, but other parasites learned to use the substances produced by the host enzymes to replace those that they lack. Therefore, three kinds of disorders regarding this system could occur.

The first kind - diseases caused by enzymes deficiencies - happen, mostly, due to inherited mutations, which can increase or decrease the enzymatic activity. When the mutation does not produce a very dangerous change in enzyme activity, the diet might have an essential role in triggering the disease. Once the cause and its severity level is understood, treatments, consisting in inhibition of some other enzymes whose reactants are less dangerous, than those accumulated due to the malfunctioning of the enzyme causing the disorder and diets that can milder it can be applied. Unfortunately, there are such diseases (see Lesch-Nyhan syndrome) where also behavioural ailment occurs, which can not be cured.

Until now, each disease triggered by an enzyme deficiency was studied after the occurrence of a patient. An analysis based on elementary flux modes could provide important information for preventing the enzyme deficiency diseases before their appearance, and also for detecting and curing them if they are already manifested. It could propose several clinical additional laboratory tests. The concentrations of several enzymes and their structure provide information about the capabilities of those enzyme to function well. The concentrations of several intermediary substances enlighten the malfunctioning of the enzymes that are supposed to degrade those substrates. The

toxicity degree of several intermediaries could explain several symptoms. A good treatment could consist in the inhibition of an enzyme at a previous step in the modes blocked by the malfunctioning of a deficient enzyme. The toxicity of the reactants that are supposed to accumulate has to be tested. Also the possibility to construct artificial degradation paths can be pointed out by such an analysis.

A wide knowledge of nucleotide metabolism may help not only to prevent, detect, or cure enzymes deficiency diseases, but also to fight against proliferative and autoimmune diseases (the second kind) or against those disorders caused by viral or parasites infections (the third kind). Without going deeply into detail, we have presented several examples gathered from the medical literature, describing various attempts to cure such ailments. We used the pathway analysis on purine and pyrimidine metabolism to support or even to extend the set of enzymes that could represent effective targets for chemotherapy. This is to indicate potential fields of application of metabolic pathway analysis in the future.

Importantly, we consider that as long as no method exist to differentiate the enzymes of the damaged cells, viruses or parasites from the enzymes of the "host", the approach of knocking out "hijacked enzymes" is not effective. Promising is to detect specific enzymes of the aggressor, essential for its replication, camouflaging, and assembling (for viruses) and inhibit them. Also creation of non-virulent mutants but easily recognisable by the immune system of the host is a nice solution and starting from pathway analysis, one can spare many efforts.

The model developed by [CCV$^+$00] is a nice example of application of pathway analysis in the fight against cancer. If it is not yet elucidated how to stop this disease, at least one can prevent additional feeding of tumour cell. Thiamine (vitamine B$_1$) is an activator of transketolase, an important enzyme in pentose phosphate pathway. The more thiamine is administered, the greater is the ribose production. This means that the purine pathway is supplied with its principal substance necessary in nucleotide production. Thus, the material required by replication can be easily provided. The avoidance to administer thiamine is a natural remedial action. Certainly, it is valuable to apply pathway analysis also in these systems that are more closely involved in DNA and RNA synthesis such as nucleotide metabolism. In purine central metabolism described in Chapter 3, the enzymes having a high degree of participation in the elementary flux modes are the most suitable for being drug targets. Importantly, efforts have to be developed to restrict the drug effect only on the cancerous cells.

6.3   Petri nets in biochemical systems analysis

Petri nets provide a special formalism to describe processes in networks. In particular, they are suitable to model biochemical networks. Here, we have shown that several concepts from Petri net theory have a significance for this modelling. However, there are alternative formalisms, and it is difficult to decide which formalism is best suited. To implement the calculations on computer, one usually translates Petri nets into matrices. So one may argue that the networks could be modelled by matrices from the very beginning. Indeed, Petri nets have the advantage to provide a means of visualisation. On the other hand, biochemists use a special way of visualisation for decades [Str95, KG91](and chemists already for centuries). Multimolecular reactions such as "A+B gives C+D+E" are represented by an arrow that has two upper ends and three lower ends. This arrow can be represented, in a formal language, as a pair of n-tuples: ((A,B), (C,D,E)). In contrast, in a normal graph as used in graph theory, edges correspond to simple pairs of nodes. In Petri nets, the representation is "disentangled" by introducing additional nodes and arcs. The above reaction would then be represented by five place nodes and one transition node (T) linked by five arcs. The arcs correspond to the following pairs of nodes: (A,T), (B,T), (T,C), (T,D) and (T,E). It is a matter of taste which representation is preferred – one pair of n-tuples or several pairs.

Many concepts from Petri net theory have counterparts in traditional biochemical modelling, for example, P-invariants (conservation relations), T-invariants (flux modes), and minimal T-invariants (elementary flux modes). In metabolism, minimal T-invariants can be interpreted as biochemical pathways. Detection of these in complex networks is often not straightforward. It is helpful in determining maximal conversion yields [RB01, SFD00, VDL02].

The concepts of trap, siphon, deadlock, and liveness, among others, have not been considered in biochemical modelling so far. Here, we have shown that these are helpful to characterize special properties of metabolic networks. For example, the test for deadlock-freeness helps to determine whether a biochemical pathway can attain a false equilibrium, where it is blocked. From another point of view, this situation has been referred to as the danger of a turbo design of pathways [TWvDW98]. The liveness of a system indicates that all transitions are able to fire infinitely often, and the processes are not eventually restricted to a subsystem. Traps can correspond to storage metabolites that are produced during growth of an organism and steadily increase in their concentrations.

We have here analysed the example of the energy metabolism in *T. brucei*. If the accumulations in the trap exceed a certain amount, this can cause product inhibition of some transitions (the aldolase reaction in the example), forcing the system to stop working. This result is of interest for elementary-modes analysis. It has been argued that this analysis can help assert the effects of enzyme deficiencies and knockout mutations [KS98, SFD00]. The example analysed here shows that in a deficient system, the remaining elementary modes may not be functional because of occurrence of a trap. Therefore, pathway analysis should be refined by considering traps, siphons, deadlock-freeness and liveness.

Siphons can correspond to storage substances when they are gradually depleted during starvation. An analysis of traps and siphons appears to be promising in studying diseases such as obesity and hypercholesterolemia, which are related to over-accumulation of storage substances. It will be worth including the analysis of traps, siphons, deadlocks, and liveness in metabolic simulation packages.

So far, in Petri net theory, transitions are always considered to be unidirectional. However, many biochemical reactions such as all isomerases are known to be reversible in that their net flow can change sign depending on the physiological state. If such a reaction is described by two oppositely directed transitions, meaningless T-invariants arise. For example, in the scheme shown in Fig. 4.3, the T-invariant {T1, T2} occurs. In order to avoid the cancellation of such T-invariants after their computation, it will be worthwhile extending Petri net theory by allowing for reversible transitions.

A property that can be checked for Petri nets is boundedness. As biochemical networks are open systems, they are not usually covered by semi-positive P-invariants; that is, they are not conservative. Nevertheless, subnets are often covered by such invariants and are, therefore, bounded. For example, if the conservation relation ATP + ADP = const. holds, one can deduce that the energy currency metabolite ATP cannot exceed a certain limit. If negative coefficients exist in the conservation relation, boundedness cannot be guaranteed even for the corresponding subnet. Beside conservative subnets, there may be superconservative subnets. Obviously, they imply unboundedness. First, there may be metabolites that are only produced by irreversible reactions but not consumed by any reaction (Fig. 4.5). Second, if consuming reactions exist, the catalysing enzymes may have such a low maximal velocity (saturation level) that the rate of production is higher than the rate of consumption.

In the third chapter (3), we have kept focussing on topological analysis, which deals with the properties that occur from the static construction of the network. For many biological applications, such as the assign-

ment of the metabolic function to an enzyme gene (functional genomics) [SBG⁺96, BDDL⁺98, DSS⁺99, FGN02], it is sufficient to analyse these properties rather than the dynamics. The structural properties are the most representative features that one should look for. Compared to kinetic parameters of enzymes, they are constant in time and often much better known. Thus, reaction stoichiometries are easier to get from databases [SMO⁺97, KG91]. Topological analysis, (in particular, the computation of invariants) constitutes the basis for the simulation of the dynamics of the system.

6.4 Routes in signalling networks

For studying signalling maps, like in the case of metabolic networks, two approaches are appropriate. One is based on differential equations and the other, on a topological analysis. The first one might solve the problem very elegantly, but has also the major inconvenience that the rate laws and kinetic data are not always known. The alternative has the advantage that it needs only stoichiometric features, which are available in many cases. However, this type of analysis leads to much more restricted conclusions. Our aim is to decompose very complicated networks into simpler parts, so that it becomes clear which factors are influencing with target, on which ways, and with which final effects. The robustness or the vulnerability of a network can be evaluated by counting the alternatives routes acting on the same targets. The idea is somehow similar to that of elementary flux modes for metabolic systems, for the identification of which there are already several programs. One of them is METATOOL [PSVNn⁺99]. Although many similarities can be found between metabolic and signalling networks, the existence of some differences does not allow us to simply apply the concept of elementary modes. Let us recall only the usual absence of mass flow conservation, the often discontinuous reaction flux and the non-stationary dynamics.

Here, we have introduced several terms, such as balanced route, structurally balanced route, balanced node, and simple route, to provide a conceptual and mathematical basis for analysing the topological structure of signalling networks. Using these concepts, we have devised an algorithm for computing all simplest balanced, cyclic and conflict routes, and the corresponding sufficient initial and the resulting final configurations in such networks. We have also developed a computer program, written in C++ and called SigNetRouter, performing that algorithm. It is easy to use and can be included as a subroutine into larger simulation packages.

Since we have adopted a Petri net formalism, the amounts of substances must be integer numbers (tokens). When the places in the net stand for

chemicals (proteins, hormones, etc), this is not a problem, all the more because the molecule numbers per substance species are usually small in signal transduction. An extension will be interesting if factors would be allowed to be environmental features, for example temperature, mechanical vibrations or osmotic stress. In addition, such factors affect each reaction (or effect) individually. An idea could be to give another interpretation to tokens for such special factors. A defined number of tokens could mean that the specified factor increases, while another number could mean that it decreases. The magnitude of increase or decrease cannot, however, be quantified in this way.

We assume that the reactions and effects that are involved in the investigated networks are irreversible. This is not a severe limitation, since signalling pathway reactions are normally irreversible because the reverse processes are catalysed by different enzymes. The signal amplification process is often made by phosphorylations, catalysed by protein kinases. In turn, the signal termination process by dephosphorylation involves many protein phosphatases.

Our model gives information only regarding the structural properties. It yields the possible routes, which the system can be decomposed into. These routes are complete to the extent to which the map is complete. Because, in signalling cascades, there is no mass flux condition, each simple route built from extending a simplest route by balancing intermediaries is also representative for signal sending. Thus, if some internal conditions prevent a given reaction or activation from proceeding, in all routes containing it, we have to cancel the transition in question and the branch following it. It is important that we predict all the possible routes. Whether they are really acting is a problem of time conditions (for example, synchrony of events), which are not regarded in our model.

Genetic engineering can be used to over-express or knock out some enzymes. This can increase or decrease, respectively, the effect of the route(s) involving them. This could be a way to control the cell's response to some specific factors and to treat dangerous diseases such as cancer [TB01].

If a route can be represented only as a sequence of monomolecular activations, inhibitions and reactions, we can evaluate its total effect (negative or positive) on the targets, multiplying the partial effects. Inhibiting a substance that inhibits another substance, it is in fact an activation. But if the route involves bi- or multimolecular reactions, only qualitative knowledge is not enough to estimate the total effect. Therefore, we have, in this case, excluded networks containing inhibitions. In future studies, it is of course worthwhile extending the analysis to this type of networks. A first step would be to allow for networks in which, although multimolecular reactions and/or

effects occur, some routes only involve monomolecular reactions, activations and inhibitions. Actually, the B cell antigen receptor signalling network used here as a biological example is of this type. However, excluding the inhibitions from signalling networks could lead to totally wrong results. See the example presented in the previous section, where one could wrongly say that the networks is not robust, just because the route containing inhibition were removed.

For the B cell antigen receptor signalling network, we have obtained, with the program SigNetRouter, 21 balanced routes. This extends results by [Cam99]. Some of the routes operate in parallel and, hence, the system is structurally quite robust.

In signalling maps such as the MAPK cascade, there are often two phosphorylations of the same protein, one after another. Usually only the last form helps in a further activation. Therefore, in our model only the last active state is marked with grey tokens. Moreover, the intermediary step is not any more considered. Of course, if also the intermediary active form will serve further for activation, it has to be somehow included in model. An appropriate solution will be to treat each intermediary substrate as a different substrate.

Appendix A

Enzymes acting in central purine metabolism depicted in Fig. 3.2.

EC number	Enzyme name	Enzyme abbreviation	Reactants	Products	Directionality
1.1.1.204	xanthine dehydrogenase	Xd	Xanthine + NAD + H2O	urate + NADH2	i
1.1.1.204			Hypoxanthine + NAD + H2O	Xanthine + NADH2	i
1.1.1.205	IMP dehydrogenase	IMPd	IMP + NAD + H2O	XMP + NADH2	i
1.1.3.22	xanthine oxidase	XOR	Xanthine + H2O + O2	urate + H2O2	i
1.17.4.1	ribonucleoside-diphosphate reductase	RNDPr	ADP + rT	dADP + oT + H2O	i
1.17.4.1			GDP + rT	dGDP + oT + H2O	i
1.7.1.7	GMP reductase	GMPr	GMP + NADPH + H+	IMP + NH3 + NADP+	i
2.1.2.2	GAR transformylase	GARt	FTH + GAR	TH + FGAR	i
2.1.2.3	AICAR transformylase	AICARt	FTH + AICAR	TF + FAICAIR	i
2.4.2.14	amidophosphoribosyltransferase	amidoPRt	PRA + PPi+ glutamate	glutamine + PRPP+ H2O	i
2.7.6.1	phosphoribosyl pyrophosphate synthetase	PRPPs	ATP + R5P	AMP + PRPP	i

EC number	Enzyme name	Enzyme abbreviation	Reactants	Products	Directionality
2.4.2.1	Purine-nucleoside phosphorylase	PNPase	dInosine + dR1P	Hypoxanthine + Pi	i
2.4.2.1			Adenine + R1P	dAdenosine + Pi	r
2.4.2.1			Hypoxanthine + R1P	Inosine + Pi	r
2.4.2.1			Adenosine + Pi	Adenine + R1P	r
2.4.2.1			Xanthosine + Pi	Xanthine + R1P	r
2.4.2.1			Guanosine + Pi	Guanine + R1P	r
2.4.2.1			Guanine + dR1P	dGuanosine + Pi	r
2.4.2.4	thymidine phosphorylase	TPase	dInosine + dR1P	Hypoxanthine + Pi	i
2.4.2.7	adenine phosphoribosyltransferase	APRT	AMP + PPi	Adenine + PRPP	r
2.4.2.7			GMP + PPi	Guanine + PRPP	r
2.4.2.8	Hypoxanthine phosphoribosyltransferase	HPRT	AMP + PPi	Adenine + PRPP	r
2.4.2.8			XMP + PPi	Xanthine + PRPP	r
2.4.2.8			GMP + PPi	Guanine + PRPP	r
2.4.2.8			IMP + PPi	Hypoxanthine + PRPP	r

EC number	Enzyme name	Enzyme abbreviation	Reactants	Products	Directionality
2.7.1.20	adenosine kinase	ADK	ATP + Adenosine	ADP + AMP	r
2.7.1.40	pyruvate kinase	KPY	ATP + pyruvate	ADP + PEP	r
2.7.1.40			dATP + pyruvate	dADP + PEP	r
2.7.1.40			GTP + pyruvate	GDP + PEP	r
2.7.1.40			dGTP + pyruvate	dGDP + PEP	r
2.7.1.74	deoxycytidine kinase	dKCY	NTP + dCytidine	NDP + dCMP	r
2.7.1.113	deoxyguanosine kinase	dKGU	ATP + dGuanosine	ADP + dGMP	r
2.7.4.3	adenylate kinase	KAD	ATP + AMP	2 ADP	r
2.7.4.3			ATP + dAMP	ADP + dADP	r
2.7.4.6	Nucleoside-diphosphate kinase	NDK	ATP + dADP	ADP + dATP	r
2.7.4.6			ATP + IDP	ADP + ITP	r
2.7.4.6			ATP + GDP	ADP + GTP	r
2.7.4.6			ATP + dGDP	ADP + dGTP	r
2.7.4.6			ATP + dIDP	ADP + dITP	r
2.7.4.8	guanylate kinase	GUK	ATP + GMP	ADP + GDP	r
2.7.4.8			ATP + dGMP	ADP + dGDP	r
2.7.7.6	RNA polymerase	RNAp	ATP + RNAn	PPi + RNAn+1	r
2.7.7.6			GTP + RNAn	PPi + RNAn+1	r

EC number	Enzyme name	Enzyme abbreviation	Reactants	Products	Directionality
2.7.7.7	DNA-directed DNA polymerase	DNAp	dATP + DNAn	PPi +DNAn+1	r
2.7.7.7			dGTP + DNAn	PPi + DNAn+1	r
3.1.3.5	AMPase	AMPase	dGMP + H2O	dGuanosine + Pi	r
3.1.3.5			XMP + H2O	Xanthosine + Pi	r
3.1.3.5			AMP + H2O	Adenosine + Pi	r
3.1.3.5			dAMP + H2O	dAdenosine + Pi	r
3.1.3.5			IMP + H2O	Inoside + Pi	r
3.1.3.5			GMP + H2O	Guanosine + Pi	r
3.1.4.17	cyclic nucleotide phosphodiesterase	cNPDe	cGMP + H2O	GMP	i
3.1.4.17			cAMP+ H2O	AMP	i
3.5.4.3	guanine deaminase	GDA	Guanine + H2O	Xanthine + NH3	i
3.5.4.4	adenosine deaminase	ADA	Adenosine + H2O	Inosine + NH3	r
3.5.4.4			dAdenosine + H2O	dInosine + NH3	r
3.5.4.6	AMP deaminase	AMPda	AMP + H2O	IMP + NH3	i
3.5.4.10	IMP cyclohydrolasee	IMPc	IMP + H2O	FAICAIR	r
3.6.1.3	adenosinetriphosphatase	ATPase	ATP + H2O	ADP + Pi	i

EC number	Enzyme name	Enzyme abbreviation	Reactants	Products	Directionality
3.6.1.5	apyrase	ADPase	GTP + H2O	GDP + Pi	i
3.6.1.5			GDP + H2O	GMP + Pi	i
3.6.1.5			ITP + H2O	IDP + Pi	i
3.6.1.5			IDP + H2O	IMP + Pi	i
3.6.1.5			ATP + H2O	ADP + Pi	i
3.6.1.5			ADP + H2O	AMP + Pi	i
3.6.1.6	nucleoside-diphosphatase	GDPase	GDP + H2O	GMP + Pi	i
3.6.1.6		IDPase	IDM + H2O	IMP+ Pi	i
3.6.1.13	Adenosine diphosphoribose pyrophosphatase	ADPRPP	ADPribose + H2O	AMP + R5P	i
3.6.1.17	dinucleoside tetraphosphatase	DNTPase	GppppG + H2O	GTP + GMP	i
3.6.1.17			XppppX + H2O	XTP + GMP	i
3.6.1.19	nucleoside-triphosphate diphosphatase	NTDPase	ITP +H2O	IMP + PPi	i
3.6.1.19			XTP + H2O	XMP + PPi	i
3.6.1.19			GTP + H2O	GMP + PPi	i
3.6.1.19			dGTP + H2O	dGMP + PPi	i
4.1.1.21	phosphoribosyl-aminoimidazole carboxylase	PRAIC	CAIR	AIR + CO2	r

EC number	Enzyme name	Enzyme abbreviation	Reactants	Products	Directionality
4.3.2.2	adenylosuccinase	AIRs	NDCE-AMP	fumarate + AMP	i
4.3.2.2			SACAIR	fumarate + AMP	r
4.6.1.1	adenylate cyclase	ADc	ATP	cAMP + PPi	i
4.6.1.2	guanylate cyclase	GUc	GTP	cGMP + PPi	i
6.3.2.6	phosphoribosyl-aminoimidazole-succinocarboxamide synthase	SACAIRs	ATP + CAIR + aspartate	ADP + Pi + SACAIR	r
6.3.3.1	phosphoribosyl-aminoimidazole synthetase	PRAIS	ATP + FGAM	ADP +Pi + AIR	i
6.3.4.4	adenylosuccinate synthase	ASS	GTP + IMP +aspartate	GDP + Pi + NDCE-AMP	i
6.3.4.13	phosphoribosyl-glycinamide synthetase	PRGAS	ATP + PRA +glycine	ADP + Pi + GAR	i
6.3.5.2	GMP synthase (glutamine-hydrolysing)	GMPs	ATP + XMP + glutamine + H2O	AMP+ PPi + GMP + glutamate	i
6.3.5.3	Phosphoribosyl-formylglycinamidine synthase	FGAMs	ATP + FGAR +glutamine + H2O	AMP + PPi + FGAM + glutamate	i

Appendix B

Metatool input file representing the system depicted in Fig. 3.2.

-ENZREV

PNPase2 PNPase3 PNPase5 PNPase7 ADK KAD NDK3 GUK AMPase AM-
Pase3 AMPase4 AMPase6 IMPc ADA AIRc AIRs SACAIRs

-ENZIRR

Xd Xd2 IMPd XOR GMPr GARt AICARt amidoPRt PRPPs cNPDe cNPDe2
KPY GDA AMPda NTDPase3 ATPase ADPase2 ADPase7 GDPase AIRs2 ADc
GUc AIRc2 GARs ASS GMPs FGAMs APRT HPRT2 HPRT3 HPRT4

-METINT

Adenine Adenosine Xanthine Xanthosine Hypoxanthine Inosine Guanine Guano-
sine R1P PRA PRPP NDCE_AMP cGMP cAMP GAR FGAR AICAR FAICAIR
SACAIR FGAM AIR CAIR

-METEXT

NAD urate NADH2 H2O2 NADPH NH3 Pi PPi ADP ATP fumarate glycine
aspartate glutamine glutamate pyruvate PEP NADP R5P GTP GDP XMP IMP
AMP GMP

-CAT

Xd : Xanthine + NAD = urate + NADH2 .

Xd2 : Hypoxanthine + NAD = Xanthine + NADH2 .

IMPd : IMP + NAD = XMP + NADH2 .

XOR : Xanthine = urate + H2O2 .

GMPr : GMP + NADPH = IMP + NH3 + NADP .

GARt : GAR = FGAR .

AICARt : AICAR = FAICAIR .

amidoPRt : glutamine + PRPP = PRA + PPi + glutamate .

PRPPs : ATP + R5P = AMP + PRPP .

cNPDe : cGMP = GMP .

cNPDe2 : cAMP = AMP .

GDA : Guanine = Xanthine + NH3 .

AMPda : AMP = IMP + NH3 .

NTDPase3 : GTP = GMP + PPi .

ATPase : ATP = ADP + Pi .

ADPase2 : GTP = GDP + Pi .

ADPase7 : ADP = AMP + Pi .

GDPase : GDP = GMP + Pi .

AIRs : SACAIR = fumarate + AICAR .

ADc : ATP = cAMP + PPi .

GUc : GTP = cGMP + PPi .

AIRc2 : ATP + FGAM = ADP + Pi + AIR .

GARs : ATP + PRA + glycine = ADP + Pi + GAR .

ASS : GTP + IMP + aspartate = GDP + Pi + NDCE_AMP .

GMPs : ATP + XMP + glutamine = AMP + PPi + GMP + glutamate .

FGAMs : ATP + FGAR + glutamine = AMP + PPi + FGAM + glutamate .

PNPase2 : Hypoxanthine + R1P = Inosine + Pi .

PNPase3 : Adenosine + Pi = Adenine + R1P .
PNPase5 : Xanthosine + Pi = Xanthine + R1P .
PNPase7 : Guanine + R1P = Guanosine + Pi .
APRT : Adenine + PRPP = AMP + PPi .
HPRT2 : Xanthine + PRPP = XMP + PPi .
HPRT3 : Guanine + PRPP = GMP + PPi .
HPRT4 : Hypoxanthine + PRPP = IMP + PPi .
ADK : ATP + Adenosine = ADP + AMP .
KPY : ADP + PEP = ATP + pyruvate .
KAD : ATP + AMP = 2 ADP .
NDK3 : ATP + GDP = ADP + GTP .
GUK : ATP + GMP = ADP + GDP .
AMPase : AMP = Adenosine + Pi .
AMPase3 : IMP = Inosine + Pi .
AMPase4 : GMP = Guanosine + Pi .
AMPase6 : XMP = Xanthosine + Pi .
IMPc : FAICAIR = IMP .
ADA : Adenosine = Inosine + NH3 .
AIRc : AIR = CAIR .
AIRs2 : NDCE_AMP = fumarate + AMP .
SACAIRs : ATP + CAIR + aspartate = ADP + Pi + SACAIR .

Appendix C

Elementary flux modes in the purine synthesis system depicted in Fig. **3.2** and their significance.

No.	Enzymes set	Overall reaction	Significance	Directionality
1	KAD	ATP + AMP = 2 ADP	Nucleotide conversion	reversible
2	NDK3	ATP + GDP = ADP + GTP	Nucleotide conversion	reversible
3	GUK	ATP + GMP = ADP + GDP	Nucleotide conversion	reversible
4	IMPd	NAD + IMP = NADH2 + XMP	Nucleotide conversion	irreversible
5	GMPr	NADPH + GMP = NH3 + NADP + IMP	Nucleotide conversion	irreversible
6	cNPDe GUc	GTP = PPi + GMP	Nucleotide conversion	irreversible
7	cNPDe2 ADc	ATP = PPi + AMP	Nucleotide conversion	irreversible
8	KPY	ADP + PEP = ATP + pyruvate	Nucleotide conversion	irreversible
9	AMPda	AMP = NH3 + IMP	Nucleotide conversion	irreversible
10	NTDPase3	GTP = PPi + GMP	Nucleotide conversion	irreversible
11	ATPase	ATP = Pi + ADP	Nucleotide conversion	irreversible
12	ADPase2	GTP = Pi + GDP	Nucleotide conversion	irreversible

No.	Enzymes set	Overall reaction	Significance	Directionality
13	ADPase7	ADP = Pi + AMP	Nucleotide conversion	irreversible
14	GDPase	GDP = Pi + GMP	Nucleotide conversion	irreversible
15	As2 ASS	aspartate + GTP + IMP = Pi + fumarate + GDP + AMP	Nucleotide conversion	irreversible
16	GMPs	ATP + glutamine + XMP = PPi + glutamate + AMP + GMP	Nucleotide conversion	irreversible
17	ADK AMPase	ATP = Pi + ADP	Nucleotide conversion	reversible
18	ADK AMPase3 -ADA	NH3 + ATP + IMP = Pi + ADP + AMP	Nucleotide conversion	reversible
19	AMPase -AMPase3 ADA	AMP = NH3 + IMP	Nucleotide conversion	reversible
20	-PNPase5 -PNPase7 AMPase4 - AMPase6 GDA	GMP = NH3 + XMP	Nucleotide conversion	irreversible
21	-PNPase2 PNPase5 AMPase3 AMPase6 (2 Xd) Xd2	3 NAD + XMP + IMP = 2 urate + 3 NADH2 + 2 Pi	Degradation to urate	irreversible
22	-PNPase2 PNPase5 -ADK AMPase6 ADA (2 Xd) Xd2	3 NAD + ADP + XMP + AMP = 2 urate + 3 NADH2 + NH3 + Pi + ATP	Degradation to urate	irreversible
23	-PNPase2 PNPase5 AMPase AMPase6 ADA (2 Xd) Xd2	3 NAD + XMP + AMP = 2 urate + 3 NADH2 + NH3 + 2 Pi	Degradation to urate	irreversible

No.	Enzymes set	Overall reaction	Significance	Directionality
24	-PNPase2 PNPase5 AMPase3 AM-Pase6 Xd2 (2 XORR)	NAD + XMP + IMP = 2 urate + NADH2 + 2 H2O2 + 2 Pi	Degradation to urate	irreversible
25	-PNPase2 PNPase5 -ADK AM-Pase6 ADA Xd2 (2 XORR)	NAD + ADP + XMP + AMP = 2 urate + NADH2 + 2 H2O2 + NH3 + Pi + ATP	Degradation to urate	irreversible
26	-PNPase2 PNPase5 AMPase AM-Pase6 ADA Xd2 (2 XORR)	NAD + XMP + AMP = 2 urate + NADH2 + 2 H2O2 + NH3 + 2 Pi	Degradation to urate	irreversible
27	-PNPase2 -PNPase7 AMPase3 AM-Pase4 (2 Xd) Xd2 GDA	3 NAD + IMP + GMP = 2 urate + 3 NADH2 + NH3 + 2 Pi	Degradation to urate	irreversible
28	-PNPase2 -PNPase7 -ADK AM-Pase4 ADA (2 Xd) Xd2 GDA	3 NAD + ADP + AMP + GMP = 2 urate + 3 NADH2 + 2 NH3 + Pi + ATP	Degradation to urate	irreversible
29	-PNPase2 -PNPase7 AMPase AM-Pase4 ADA (2 Xd) Xd2 GDA	3 NAD + AMP + GMP = 2 urate + 3 NADH2 + 2 NH3 + 2 Pi	Degradation to urate	irreversible
30	-PNPase2 -PNPase7 AMPase3 AM-Pase4 Xd2 (2 XORR) GDA	NAD + IMP + GMP = 2 urate + NADH2 + 2 H2O2 + NH3 + 2 Pi	Degradation to urate	irreversible
31	-PNPase2 -PNPase7 -ADK AM-Pase4 ADA Xd2 (2 XORR) GDA	NAD + ADP + AMP + GMP = 2 urate + NADH2 + 2 H2O2 + 2 NH3 + Pi + ATP	Degradation to urate	irreversible
32	-PNPase2 -PNPase7 AMPase AM-Pase4 ADA Xd2 (2 XORR) GDA	NAD + AMP + GMP = 2 urate + NADH2 + 2 H2O2 + 2 NH3 + 2 Pi	Degradation to urate	irreversible

No.	Enzymes set	Overall reaction	Significance	Directionality
33	IMPc Ac As SACAs GARt AICARt amidoPRt PRPPs Ac2 GARs FGAMs	5 ATP + glycine + aspartate + 2 glutamine + R5P = 3 Pi + 2 PPi + 3 ADP + fumarate + 2 glutamate + IMP + 2 AMP	De novo synthesis	irreversible
34	-PNPase2 PNPase3 -ADK AMPase3 (2 PRPPs) APRT HPRT4	ADP + ATP + 2 R5P = Pi + 2 PPi + 2 AMP	Nucleotide conversion	irreversible
35*	-PNPase2 (2 PNPase3) -PNPase5 (-2 ADK) AMPase3 -AMPase6 Xd2 (2 PRPPs) (2 APRT)	NAD + 2 ADP + 2 R5P + IMP = NADH2 + 2 PPi + XMP + 2 AMP	Nucleotide conversion	irreversible
36	-PNPase2 PNPase3 -ADK AMPase3 Xd Xd2 PRPPs APRT	2 NAD + ADP + R5P + IMP = urate + 2 NADH2 + Pi + PPi + AMP	Degradation to urate	irreversible
37	-PNPase2 PNPase3 -ADK AMPase3 Xd2 XORR PRPPs APRT	NAD + ADP + R5P + IMP = urate + NADH2 + H2O2 + Pi + PPi + AMP	Degradation to urate	irreversible
38*	-PNPase2 PNPase3 -ADK AMPase3 Xd2 (2 PRPPs) APRT HPRT2	NAD + ADP + ATP + 2 R5P + IMP = NADH2 + Pi + 2 PPi + XMP + 2 AMP	Nucleotide conversion	irreversible
39	-PNPase2 PNPase3 (-2 ADK) ADA (2 PRPPs) APRT HPRT4	2 ADP + 2 R5P = NH3 + 2 PPi + IMP + AMP	Nucleotide conversion	irreversible
40*	-PNPase2 (2 PNPase3) -PNPase5 (-3 ADK) -AMPase6 ADA Xd2 (2 PRPPs) (2 APRT)	NAD + Pi + 3 ADP + 2 R5P = NADH2 + NH3 + 2 PPi + ATP + XMP + AMP	Nucleotide conversion	irreversible

* Alternative elementary flux modes capable to produce XMP

No.	Enzymes set	Overall reaction	Significance	Directionality
41	-PNPase2 PNPase3 (-2 ADK) ADA Xd Xd2 PRPPs APRT	2 NAD + 2 ADP + R5P = urate + 2 NADH2 + NH3 + PPi + ATP	Degradation to urate	irreversible
42	-PNPase2 PNPase3 (-2 ADK) ADA Xd2 XORR PRPPs APRT	NAD + 2 ADP + R5P = urate + NADH2 + H2O2 + NH3 + PPi + ATP	Degradation to urate	irreversible
43*	-PNPase2 PNPase3 (-2 ADK) ADA Xd2 (2 PRPPs) APRT HPRT2	NAD + 2 ADP + 2 R5P = NADH2 + NH3 + 2 PPi + XMP + AMP	Nucleotide conversion	irreversible
44	-PNPase2 PNPase3 AMPase AMPase3 (2 PRPPs) APRT HPRT4	ATP + R5P = Pi + PPi + AMP	Nucleotide conversion	irreversible
45*	-PNPase2 (2 PNPase3) -PNPase5 (2 AMPase) AMPase3 -AMPase6 Xd2 (2 PRPPs) (2 APRT)	NAD + 2 ATP + 2 R5P + IMP = NADH2 + 2 Pi + 2 PPi + XMP + 2 AMP	Nucleotide conversion	irreversible
46	-PNPase2 PNPase3 AMPase AMPase3 Xd Xd2 PRPPs APRT	2 NAD + ATP + R5P + IMP = urate + 2 NADH2 + 2 Pi + PPi + AMP	Degradation to urate	irreversible
47	-PNPase2 PNPase3 AMPase AMPase3 Xd2 XORR PRPPs APRT	NAD + ATP + R5P + IMP = urate + NADH2 + H2O2 + 2 Pi + PPi + AMP	Degradation to urate	irreversible
48*	-PNPase2 PNPase3 AMPase AMPase3 Xd2 (2 PRPPs) APRT HPRT2	NAD + 2 ATP + 2 R5P + IMP = NADH2 + 2 Pi + 2 PPi + XMP + 2 AMP	Nucleotide conversion	irreversible
49	-PNPase2 PNPase3 (2 AMPase) ADA (2 PRPPs) APRT HPRT4	2 ATP + 2 R5P = NH3 + 2 Pi + 2 PPi + IMP + AMP	Nucleotide conversion	irreversible

* Alternative elementary flux modes capable to produce XMP

No.	Enzymes set	Overall reaction	Significance	Directionality
50*	-PNPase2 (2 PNPase3) -PNPase5 (3 AMPase) -AMPase6 ADA Xd2 (2 PRPPs) (2 APRT)	NAD + 2 ATP + 2 R5P = NADH2 + NH3 + 2 Pi + 2 PPi + XMP + AMP	Nucleotide conversion	irreversible
51	-PNPase2 PNPase3 (2 AMPase) ADA Xd Xd2 PRPPs APRT	2 NAD + ATP + R5P = urate + 2 NADH2 + NH3 + 2 Pi + PPi	Degradation to urate	irreversible
52	-PNPase2 PNPase3 (2 AMPase) ADA Xd2 XORR PRPPs APRT	NAD + ATP + R5P = urate + NADH2 + H2O2 + NH3 + 2 Pi + PPi	Degradation to urate	irreversible
53*	-PNPase2 PNPase3 (2 AMPase) ADA Xd2 (2 PRPPs) APRT HPRT2	NAD + 2 ATP + 2 R5P = NADH2 + NH3 + 2 Pi + 2 PPi + XMP + AMP	Nucleotide conversion	irreversible
54	-PNPase2 PNPase5 AMPase3 AMPase6 Xd PRPPs HPRT4	NAD + ATP + R5P + XMP = urate + NADH2 + 2 Pi + PPi + AMP	Degradation to urate	irreversible
55	-PNPase2 PNPase5 -ADK AMPase6 ADA Xd PRPPs HPRT4	NAD + ADP + R5P + XMP = urate + NADH2 + NH3 + Pi + PPi + IMP	Degradation to urate	irreversible
56	-PNPase2 PNPase5 AMPase AMPase6 ADA Xd PRPPs HPRT4	NAD + ATP + R5P + XMP = urate + NADH2 + NH3 + 2 Pi + PPi + IMP	Degradation to urate	irreversible
57	-PNPase2 PNPase5 AMPase3 AMPase6 XOR PRPPs HPRT4	ATP + R5P + XMP = urate + H2O2 + 2 Pi + PPi + AMP	Degradation to urate	irreversible
58	-PNPase2 PNPase5 -ADK AMPase6 ADA XOR PRPPs HPRT4	ADP + R5P + XMP = urate + H2O2 + NH3 + Pi + PPi + IMP	Degradation to urate	irreversible

* Alternative elementary flux modes capable to produce XMP

No.	Enzymes set	Overall reaction	Significance	Directionality
59	-PNPase2 PNPase5 AMPase AMPase6 ADA XOR PRPPs HPRT4	ATP + R5P + XMP = urate + H2O2 + NH3 + 2 Pi + PPi + IMP	Degradation to urate	irreversible
60	-PNPase2 PNPase5 AMPase3 AMPase6 (2 PRPPs) HPRT4	ATP + R5P = Pi + PPi + AMP	Nucleotide conversion	irreversible
61*	-PNPase2 PNPase5 AMPase3 AMPase6 Xd2 (2 PRPPs) (2 HPRT2)	NAD + 2 ATP + 2 R5P + IMP = NADH2 + 2 Pi + 2 PPi + XMP + 2 AMP	Nucleotide conversion	irreversible
62	-PNPase2 PNPase5 -ADK AMPase6 ADA (2 PRPPs) HPRT2 HPRT4	ADP + ATP + 2 R5P = NH3 + Pi + 2 PPi + IMP + AMP	Nucleotide conversion	irreversible
63*	-PNPase2 PNPase5 -ADK AMPase6 ADA Xd2 (2 PRPPs) (2 HPRT2)	NAD + ADP + ATP + 2 R5P = NADH2 + NH3 + Pi + 2 PPi + XMP + AMP	Nucleotide conversion	irreversible
64	-PNPase2 PNPase5 AMPase AMPase6 ADA (2 PRPPs) HPRT2 HPRT4	2 ATP + 2 R5P = NH3 + 2 Pi + 2 PPi + IMP + AMP	Nucleotide conversion	irreversible
65*	-PNPase2 PNPase5 AMPase AMPase6 ADA Xd2 (2 PRPPs) (2 HPRT2)	NAD + 2 ATP + 2 R5P = NADH2 + NH3 + 2 Pi + 2 PPi + XMP + AMP	Nucleotide conversion	irreversible
66	-PNPase2 PNPase3 (2 AMPase3) -ADA (2 PRPPs) APRT HPRT4	NH3 + 2 ATP + 2 R5P + IMP = 2 Pi + 2 PPi + 3 AMP	Nucleotide conversion	irreversible
67*	-PNPase2 (2 PNPase3) -PNPase5 (3 AMPase3) -AMPase6 (-2 ADA) Xd2 (2 PRPPs) (2 APRT)	NAD + 2 NH3 + 2 ATP + 2 R5P + 3 IMP = NADH2 + 2 Pi + 2 PPi + XMP + 4 AMP	Nucleotide conversion	irreversible

* Alternative elementary flux modes capable to produce XMP

No.	Enzymes set	Overall reaction	Significance	Directionality
68	-PNPase2 PNPase3 (2 AMPase3) - ADA Xd Xd2 PRPPs APRT	2 NAD + NH3 + ATP + R5P + 2 IMP = urate + 2 NADH2 + 2 Pi + PPi + 2 AMP	Degradation to urate	irreversible
69	-PNPase2 PNPase3 (2 AMPase3) - ADA Xd2 XOR PRPPs APRT	NAD + NH3 + ATP + R5P + 2 IMP = urate + NADH2 + H2O2 + 2 Pi + PPi + 2 AMP	Degradation to urate	irreversible
70*	-PNPase2 PNPase3 (2 AMPase3) -ADA Xd2 (2 PRPPs) APRT HPRT2	NAD + NH3 + 2 ATP + 2 R5P + 2 IMP = NADH2 + 2 Pi + 2 PPi + XMP + 3 AMP	Nucleotide conversion	irreversible
71	-PNPase2 -PNPase7 AMPase3 AMPase4 (2 PRPPs) HPRT3 HPRT4	ATP + R5P = Pi + PPi + AMP	Nucleotide conversion	irreversible
72*	-PNPase2 -PNPase5 (-2 PNPase7) AMPase3 (2 AMPase4) -AMPase6 Xd2 (2 PRPPs) (2 HPRT3)	NAD + 2 ATP + 2 R5P + IMP = NADH2 + 2 Pi + 2 PPi + XMP + 2 AMP	Nucleotide conversion	irreversible
73	-PNPase2 -PNPase7 AMPase3 AMPase4 Xd Xd2 PRPPs HPRT3	2 NAD + ATP + R5P + IMP = urate + 2 NADH2 + 2 Pi + PPi + AMP	Degradation to urate	irreversible
74	-PNPase2 -PNPase7 AMPase3 AMPase4 Xd2 XOR PRPPs HPRT3	NAD + ATP + R5P + IMP = urate + NADH2 + H2O2 + 2 Pi + PPi + AMP	Degradation to urate	irreversible

* Alternative elementary flux modes capable to produce XMP

No.	Enzymes set	Overall reaction	Significance	Directionality
75	-PNPase2 -PNPase7 AMPase3 AMPase4 Xd2 (2 PRPPs) HPRT2 HPRT3	NAD + 2 ATP + 2 R5P + IMP = NADH2 + 2 Pi + 2 PPi + XMP + 2 AMP	Nucleotide conversion	irreversible
76	-PNPase2 -PNPase7 -ADK AMPase4 ADA (2 PRPPs) HPRT3 HPRT4	ADP + ATP + 2 R5P = NH3 + Pi + 2 PPi + IMP + AMP	Nucleotide conversion	irreversible
77*	-PNPase2 -PNPase5 (-2 PNPase7) -ADK (2 AMPase4) -AMPase6 ADA Xd2 (2 PRPPs) (2 HPRT3)	NAD + ADP + ATP + 2 R5P = NADH2 + NH3 + Pi + 2 PPi + XMP + AMP	Nucleotide conversion	irreversible
78	-PNPase2 -PNPase7 -ADK AMPase4 ADA Xd Xd2 PRPPs HPRT3	2 NAD + ADP + R5P = urate + 2 NADH2 + NH3 + Pi + PPi	Degradation to urate	irreversible
79	-PNPase2 -PNPase7 -ADK AMPase4 ADA Xd2 XOR PRPPs HPRT3	NAD + ADP + R5P = urate + NADH2 + H2O2 + NH3 + Pi + PPi	Degradation to urate	irreversible
80*	-PNPase2 -PNPase7 -ADK AMPase4 ADA Xd2 (2 PRPPs) HPRT2 HPRT3	NAD + ADP + ATP + 2 R5P = NADH2 + NH3 + Pi + 2 PPi + XMP + AMP	Nucleotide conversion	irreversible
81	-PNPase2 -PNPase7 AMPase AMPase4 ADA (2 PRPPs) HPRT3 HPRT4	2 ATP + 2 R5P = NH3 + 2 Pi + 2 PPi + IMP + AMP	Nucleotide conversion	irreversible
82*	-PNPase2 -PNPase5 (-2 PNPase7) AMPase (2 AMPase4) -AMPase6 ADA Xd2 (2 PRPPs) (2 HPRT3)	NAD + 2 ATP + 2 R5P = NADH2 + NH3 + 2 Pi + 2 PPi + XMP + AMP	Nucleotide conversion	irreversible

* Alternative elementary flux modes capable to produce XMP

No.	Enzymes set	Overall reaction	Significance	Directionality
83	-PNPase2 -PNPase7 AMPase AMPase4 ADA Xd Xd2 PRPPs HPRT3	$2\ NAD + ATP + R5P = urate + 2\ NADH2 + NH3 + 2\ Pi + PPi$	Degradation to urate	irreversible
84	-PNPase2 -PNPase7 AMPase AMPase4 ADA Xd2 XOR PRPPs HPRT3	$NAD + ATP + R5P = urate + NADH2 + H2O2 + NH3 + 2\ Pi + PPi$	Degradation to urate	irreversible
85	-PNPase2 -PNPase7 AMPase AMPase4 ADA Xd2 (2 PRPPs) HPRT3	$NAD + 2\ ATP + 2\ R5P = NADH2 + NH3 + 2\ Pi + 2\ PPi + XMP + AMP$	Nucleotide conversion	irreversible
86	-PNPase2 -PNPase7 AMPase3 AMPase4 Xd PRPPs GDA HPRT4	$NAD + ATP + R5P + GMP = urate + NADH2 + NH3 + 2\ Pi + PPi + AMP$	Degradation to urate	irreversible
87	-PNPase2 -PNPase7 -ADK AMPase4 ADA Xd PRPPs GDA HPRT4	$NAD + ADP + R5P + GMP = urate + NADH2 + 2\ NH3 + Pi + PPi + IMP$	Degradation to urate	irreversible
88	-PNPase2 -PNPase7 AMPase AMPase4 ADA Xd PRPPs GDA HPRT4	$NAD + ATP + R5P + GMP = urate + NADH2 + 2\ NH3 + 2\ Pi + PPi + IMP$	Degradation to urate	irreversible
89	-PNPase2 -PNPase7 AMPase3 AMPase4 XOR PRPPs GDA HPRT4	$ATP + R5P + GMP = urate + H2O2 + NH3 + 2\ Pi + PPi + AMP$	Degradation to urate	irreversible

No.	Enzymes set	Overall reaction	Significance	Directionality
90	-PNPase2 -PNPase7 -ADK AMPase4 ADA XOR PRPPs GDA HPRT4	ADP + R5P + GMP = urate + H2O2 + 2 NH3 + Pi + PPi + IMP	Degradation to urate	irreversible
91	-PNPase2 -PNPase7 AMPase4 ADA XOR PRPPs GDA HPRT4	ATP + R5P + GMP = urate + H2O2 + 2 NH3 + 2 Pi + PPi + IMP	Degradation to urate	irreversible
92*	-PNPase2 -PNPase7 AMPase3 AMPase4 (2 PRPPs) GDA HPRT2 HPRT4	2 ATP + 2 R5P + GMP = NH3 + 2 Pi + 2 PPi + XMP + 2 AMP	Nucleotide conversion	irreversible
93*	-PNPase2 -PNPase7 AMPase3 AMPase4 Xd2 (2 PRPPs) GDA (2 HPRT2)	NAD + 2 ATP + 2 R5P + IMP + GMP = NADH2 + NH3 + 2 Pi + 2 PPi + 2 XMP + 2 AMP	Nucleotide conversion	irreversible
94*	-PNPase2 -PNPase7 -ADK AMPase4 ADA (2 PRPPs) GDA HPRT2 HPRT4	ADP + ATP + 2 R5P + GMP = 2 NH3 + Pi + 2 PPi + XMP + IMP + AMP	Nucleotide conversion	irreversible
95*	-PNPase2 -PNPase7 -ADK AMPase4 ADA Xd2 (2 PRPPs) GDA (2 HPRT2)	NAD + ADP + ATP + 2 R5P + GMP = NADH2 + 2 NH3 + Pi + 2 PPi + 2 XMP + AMP	Nucleotide conversion	irreversible
96*	-PNPase2 -PNPase7 AMPase AMPase4 ADA (2 PRPPs) GDA HPRT2 HPRT4	2 ATP + 2 R5P + GMP = 2 NH3 + 2 Pi + 2 PPi + XMP + IMP + AMP	Nucleotide conversion	irreversible
97*	-PNPase2 -PNPase7 AMPase AMPase4 ADA Xd2 (2 PRPPs) GDA (2 HPRT2)	NAD + 2 ATP + 2 R5P + GMP = NADH2 + 2 NH3 + 2 Pi + 2 PPi + 2 XMP + AMP	Nucleotide conversion	irreversible

* Alternative elementary flux modes capable to produce XMP

Appendix D

Enzymes acting in the pyrimidine pathway depicted in Fig. 3.4.

E.C. number	Enzyme name	Abbreviation	Reaction
1.3.1.2 1.3.1.2	dihydropyrimidine dehydrogenase	DPYD	5,6-dihydrouracil + NADP = uracil + NADPH2 5,6-dihydrothymine + NADP = thymine + NADPH2
1.3.3.1	dihydroorotate dehydrogenase	DHODH	(S)-Dihydroorotate + O2 = orotate + H2O2
1.8.1.9 1.8.1.9	thioredoxin reductase 1	TXNRD1	thioredoxin + NADP+ = thioredoxin disulfide + NADPH + H+
1.17.4.1 1.17.4.1	ribonucleoside-diphosphate reductase alpha chain	RRM1	dUDP + Oxidized thioredoxin + H2O = Thioredoxin + UDP dCDP + Oxidized thioredoxin + H2O = Thioredoxin + CDP
2.1.1.45	thymidylate synthetase	TYMS	5,10-methylenetetrahydrofolate + dUMP = dihydrofolate + dTMP
2.1.3.2	aspartate carbamoyltransferase	CAD	carbamoyl phosphate + L-aspartate = phosphate + N-carbamoyl-L-aspartate
2.4.2.1	nucleoside phosphorylase	NP	purine nucleoside + phosphate = purine + alpha-D-ribose 1-phosphate
2.4.2.3	uridine phosphorylase	UP	uridine + phosphate = uracil + alpha-D-ribose 1-phosphate
2.4.2.4 2.4.2.4	endothelial cell growth factor 1 (platelet-derived)	ECGF1	thymidine + phosphate = thymine + 2-deoxy-D-ribose 1-phosphatedeoxyuridine + phosphate = uracil + 2-deoxy-D-ribose 1-phosphate

E.C. number	Enzyme name	Abbreviation	Reaction
2.4.2.10 4.1.1.23	uridine monophosphate synthetase (orotate phosphoribosyl transferase and orotidine-5'-decarboxylase)	UMPS	orotidine 5'-phosphate + diphosphate = orotate + 5-phospho-alpha-D-ribose 1-diphosphate orotidine 5'-phosphate = UMP + CO
2 2.7.1.21	thymidine kinase 1, soluble	TK1	ATP + thymidine = ADP + dTMPATP + deoxyuridine = ADP + dUMP
2.7.1.48	uridine monophosphate kinase	UMPK	ATP + uridine = ADP + UMPATP + cytidine = ADP + CMP
2.7.1.74	deoxycytidine kinase	DCK	NTP + deoxycytidine = NDP + dCMP
2.7.4.6	nucleoside diphosphate kinase type 6 (inhibitor of p53-induced apoptosis-alpha)	NM23-H6	ATP + nucleoside diphosphate = ADP + nucleoside triphosphate
2.7.4.9	deoxythymidylate kinase (thymidylate kinase)	DTYMK	ATP + dTMP = ADP + dTDPATP + Thymidine = ADP + dTMPATP + dUMP = ADP + dUDP
2.7.4.10	adenylate kinase 3 alpha like	AKL3L	nucleoside triphosphate + AMP = nucleoside diphosphate + ADP
2.7.4.14	UMP-CMP kinase	UMP-CMPK	ATP + (d)CMP = ADP + (d)CDPATP + UMP = ADP + UDP
2.7.7.6	polymerase (RNA) II (DNA directed) polypeptide C	POLR2C	nucleoside triphosphate + RNAn = diphosphate + RNAn+1
2.7.7.7	MAD2 mitotic arrest deficient-like 2 (yeast)	MAD2L2	deoxynucleoside triphosphate + DNAn = diphosphate + DNAn+1

E.C. number	Enzyme name	Abbreviation	Reaction
3.1.3.5	5'-nucleotidase, cytosolic II	NT5C2	A 5'-ribonucleotide + H_2O = a ribonucleoside + phosphate
3.5.1.6	ureidopropionase, beta	UPB1	3-Ureidopropionate + H_2O = β-Alanine + CO_2 + NH_3 3-Ureidoisobutyrate + H_2O = 3-Aminoisobutanoate + CO_2 + NH_3
3.5.2.2	dihydropyrimidinase	DPYS	5,6-dihydrouracil + H_2O = 3-ureidopropanoate 5,6-dihydrothymine + H_2O = 3-ureidoisobutyrate
3.5.2.3	dihydroorotate hydrolase	CAD	(S)-dihydroorotate + H_2O = N-carbamoyl-L-aspartate
3.5.4.5	cytidine deaminase	CDA	(d)cytidine + H_2O = (d)uridine + NH_3
3.5.4.12	dCMP deaminase	DCTD	dCMP + H_2O = dUMP + NH_3
3.6.1.6	3.6.1.5 Ca2+-dependent endoplasmic reticulum nucleoside diphosphatase	SHAPY	A nucleoside diphosphate + H_2O = a nucleotide + phosphate ATP + 2 H_2O = AMP + 2 phosphate
3.6.1.17	nudix (nucleoside diphosphate linked moiety X)-type motif 2	NUDT2	P1,P4-bis(5'-guanosyl) tetraphosphate + H_2O = GTP + GMP
3.6.1.19	inosine triphosphatase (nucleoside triphosphate pyrophosphatase)	ITPA	(d)UTM + H_2O = (d)UMP + diphosphate
3.6.1.23	dUTP pyrophosphatase	DUT	dUTP + H_2O = dUMP + diphosphate
6.3.4.2	CTP synthase	CTPS	ATP + UTP + NH_3 = ADP + phosphate + CTP
6.3.5.5	carbamoyl-phosphate synthase	CAD	2 ATP + L-Gln + CO_2 + H_2O = 2 ADP + phosphate + L-Glu + carbamoyl phosphate

Appendix E

Metatool input file representing the pyrimidine metabolism depicted in Fig. 3.4.

-METINT

dUracil, uracil, DHUracil, DHthymine, dThymine, thymine, thymidine, dUridine, uridine, dCytidine, cytidine, DHO, orotate, Othioredoxin, thioredoxin2S, thioredoxin, dRNPP, RNPP, CP, CA, dTMP, dTDP, dTTP, dUMP, dUDP, dUTP dCMP, dCDP, dCTP, UMP, UDP, UTP, CMP, CDP, CTP, O5P, 3-UP, UisoB

-METEXT

PRPP,Gln, Glu, ATP, ADP, AMP, PPi, CO2,Pi, H, NADP, NADPH, NADPH2, O2, H2O2, H2O, Rthioredoxin, NH3, β-alanine, AisoB, ME4HF, DHF, R1P, dR1P, aspartate

-ENZREV:

DPYD1, DPYD2, DHODH, UP, UMPS1, NM1, NM2, NM3, NM4, NM5, DTYMK1, DTYMK2, DTYMK3, UMPCMPK1, UMPCMPK2, UMPCMPK3, DPYS1, DPYS2, CAD1

-ENZIRREV:

TXNRD1, RRM1, RRM2, TYMS, CAD3, NP, ECGF1, ECGF2, UMPS2, TK1, TK2, UMPK1, UMPK2, DCK, KAD, UMPK3, NT5C1, NT5C2, NT5C3, NT5C4, UPB1, UPB2, CDA1, CDA2, DCTD, SHAPY1, SHAPY2, SHAPY3, SHAPY4, SHAPY5, SHAPY6, ITPA1, ITPA2, DUT, CTPS, CAD2

-CAT

DPYD1 : dUracil + NADP = uracil + NADPH2 .

DPYD2 : dThymine + NADP = thymine + NADPH2 .

DHODH : DHO + O2 = orotate + H2O2 .

TXNRD1 : thioredoxin + NADP+ = thioredoxin2S + NADPH + H .

RRM1 : dUDP + Othioredoxin + H2O = thioredoxin + UDP .

RRM2 : dCDP + Othioredoxin + H2O = thioredoxin + CDP .

TYMS : ME4HF + dUMP = DHF + dTMP .

CAD3 : CP + aspartate = Pi + CA .

NP : dUridine + Pi = uracil + R1P .

UP : uridine + Pi = uracil + R1P .

ECGF1 : thymidine + Pi = thymine + dR1P .

ECGF2 : deoxyuridine + Pi = uracil + dR1P .

UMPS1 : O5P + PPi = orotate + PRPP .

UMPS2 : O5P = UMP + CO2 .

TK1 : ATP + thymidine = ADP + dTMP .

TK2 : ATP + dUridine = ADP + dUMP .

UMPK1 : ATP + uridine = ADP + UMP .

UMPK2 : ATP + cytidine = ADP + CMP .

UMPK3 : ATP + dCytidine = ADP + dCMP .

DCK : ATP + dCytidine = ADP + dCMP .

NM1 : ATP + UDP = ADP + UTP .

NM2 : ATP + CDP = ADP + CTP .

NM3 : ATP + dCDP = ADP + dCTP .

NM4 : ATP + dUDP = ADP + dUTP .

NM5 : ATP + dTDP = ADP + dTTP .
DTYMK1 : ATP + dTMP = ADP + dTDP .
DTYMK2 : ATP + Thymidine = ADP + dTMP .
DTYMK3 : ATP + dUMP = ADP + dUDP .
KAD : UTP + AMP = UDP + ADP .
UMPCMPK1 : ATP + CMP = ADP + CDP .
UMPCMPK2 : ATP + dCMP = ADP + dCDP .
UMPCMPK3 : ATP + UMP = ADP + UDP .
NT5C1 : dTMP + H2O = thymidine + Pi .
NT5C2 : dCMP + H2O = dCytidine + Pi .
NT5C3 : CMP + H2O = cytidine + Pi .
NT5C4 : UMP + H2O = uridine + Pi .
UPB1 : 3-UP + H2O = β-alanine + CO2 + NH3 .
UPB2 : 3-UisoB + H2O = 3-AisoB + CO2 + NH3 .
DPYS1 : DHUracil + H2O = 3-UP .
DPYS2 : DHthymine + H2O = UisoB .
CAD1 : DHO + H2O = CA .
CDA1 : (d)cytidine + H2O = (d)uridine + NH3 .
CDA2: dCytidine + H2O = dUridine + NH3 .
DCTD : dCMP + H2O = dUMP + NH3 .
SHAPY1 : UDP + H2O = UMP + Pi .
SHAPY2 : CDP + H2O = CMP + Pi .
SHAPY3 : UTP + H2O = UDP + Pi .
SHAPY4 : CTP + H2O = CDP + Pi .
SHAPY5 : dTTP + H2O = dTDP + Pi .
SHAPY6 : dTDP + H2O = dTMP + Pi .
ITPA1 : UTM + H2O = UMP + PPi .
ITPA2 : dUTM + H2O = dUMP + PPi .
DUT : dUTP + H2O = dUMP + PPi .
CTPS : ATP + UTP + NH3 = ADP + Pi + CTP .
CAD2 : 2 ATP + Gln + CO2 + H2O = 2 ADP + phosphate + Glu + CP .

Appendix F

Elementary flux modes and enzyme subsets in the pyrimidine system depicted in Fig. 3.4.

No	Enzymes	overall reaction
1	NM5 SHAPY5 irreversible	ATP + H2O = ADP + Pi
2	DTYMK1 SHAPY6 irreversible	ATP + H2O = ADP + Pi
3	-DTYMK2 TK1 irreversible	
4	DTYMK2 NT5C1 irreversible	ATP + H2O = ADP + Pi
5	TK1 NT5C1 irreversible	ATP + H2O = ADP + Pi
6	DCK NT5C2 irreversible	ATP + H2O = ADP + Pi
7	NM4 DTYMK3 ITPA2 irreversible	2 ATP + H2O = 2 ADP + PPi
8	NM4 DTYMK3 DUT irreversible	2 ATP + H2O = 2 ADP + PPi
9	-DPYD1 DHODH UP -UMPS1 DPYS1 -CAD1 CAD3 UMPS2 NT5C4 UPB1 CAD2 irreversible	PRPP + Gln + 2 ATP + NADPH2 + O2 + 3 H2O + aspartate = Glu + 2 ADP + PPi + CO2 + 2 Pi + NADP + H2O2 + NH3 + β-alanine + R1P
10	NM1 KAD irreversible	ATP + AMP = 2 ADP
11	NM1 SHAPY3 irreversible	ATP + H2O = ADP + Pi
12	-DPYD1 -NM3 -UMPK2 DPYS1 NP NT5C2 UPB1 CDA2 SHAPY2 irreversible	2 ADP + NADPH2 + 5 H2O + dCTP = 2 ATP + CO2 + Pi + NADP + 2 NH3 + β-alanine + R1P
13	-DPYD1 -NM3 -UMPK2 DPYS1 ECGF2 NT5C2 UPB1 CDA2 SHAPY2 irreversible	2 ADP + NADPH2 + 5 H2O + dCTP = 2 ATP + CO2 + Pi + NADP + 2 NH3 + β-alanine + dR1P
14	-DPYD2-NM3 -DTYMK2 -UMPK2 DPYS2 TYMS ECGF1 UPB2 DCTD SHAPY2 irreversible	3 ADP + NADPH2 + 4 H2O + ME4HF + dCTP = 3 ATP + CO2 + NADP + 2 NH3 + AisoB + DHF + dR1P
15	-DPYD2 -NM3 -UMPK2 DPYS2 TYMS ECGF1 NT5C1 UPB2 DCTD SHAPY2 irreversible	2 ADP + NADPH2 + 5 H2O + ME4HF + dCTP = 2 ATP + CO2 + Pi + NADP + 2 NH3 + AisoB + DHF + dR1P

No	Enzymes	overall reaction
16	-DPYD2 -NM3 -DTYMK2 -UMPK2 DPYS2 TYMS ECGF1 TK2 NT5C2 UPB2 CDA2 SHAPY2 irreversible	2 ADP + NADPH2 + 5 H2O + ME4HF + dCTP = 2 ATP + CO2 + Pi + NADP + 2 NH3 + AisoB + DHF + dR1P
17	-DPYD2 -NM3 -UMPK2 DPYS2 TYMS ECGF1 TK2 NT5C1 NT5C2 UPB2 CDA2 SHAPY2 irreversible	ADP + NADPH2 + 6 H2O + ME4HF + dCTP = ATP + CO2 + 2 Pi + NADP + 2 NH3 + AisoB + DHF + dR1P
18	UMPK1 -UMPK2 UMPK3 (2 SHAPY2) irreversible	ATP + 2 H2O = ADP + 2 Pi
19	-UMPK2 UMPK3 SHAPY1 SHAPY2 irreversible	H2O = Pi
20	NM1 -UMPK2 UMPK3 SHAPY2 ITPA1 irreversible	ATP + 2 H2O = ADP + PPi + Pi
21	-DPYD1 UP -NM3 DTYMK3 UMPK1 -UMPK2 DPYS1 TXNRD1 RRM1 NT5C4 UPB1 DCTD (2 SHAPY2) irreversible	NADPH2 + 7 H2O + dCTP = CO2 + 2 Pi + H + NADPH + 2 NH3 + β-alanine + R1P
22	-DPYD1 UP -NM3 DTYMK3 -UMPK2 DPYS1 TXNRD1 RRM1 NT5C4 UPB1 DCTD SHAPY1 SHAPY2 irreversible	ADP + NADPH2 + 7 H2O + dCTP = ATP + CO2 + 2 Pi + H + NADPH + 2 NH3 + β-alanine + R1P
23	-DPYD1 UP -NM3 DTYMK3 UMPK1 -UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C2 NT5C4 UPB1 CDA2 (2 SHAPY2) irreversible	ATP + NADPH2 + 8 H2O + dCTP = ADP + CO2 + 3 Pi + H + NADPH + 2 NH3 + β-alanine + R1P
24	-DPYD1 UP -NM3 DTYMK3 -UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C2 NT5C4 UPB1 CDA2 SHAPY1 SHAPY2 irreversible	NADPH2 + 8 H2O + dCTP = CO2 + 3 Pi + H + NADPH + 2 NH3 + β-alanine + R1P
25	-DPYD1 UP NM1 -NM3 DTYMK3 -UMPK2 DPYS1 TXNRD1 RRM1 NT5C4 UPB1 DCTD SHAPY2 ITPA1 irreversible	NADPH2 + 7 H2O + dCTP = PPi + CO2 + Pi + H + NADPH + 2 NH3 + β-alanine + R1P

No	Enzymes	overall reaction
26	-DPYD1 UP NM1 -NM3 DTYMK3 -UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C2 NT5C4 UPB1 CDA2 SHAPY2 ITPA1 irreversible	ATP + NADPH2 + 8 H2O + dCTP = ADP + PPi + CO2 + 2 Pi + H + NADPH + 2 NH3 + β-alanine + R1P
27	-DPYD1 -NM3 -UMPK1 -UMPK2 DPYS1 NP NT5C2 UPB1 CDA2 SHAPY1 irreversible	3 ADP + NADPH2 + 5 H2O + dCTP = 3 ATP + CO2 + Pi + NADP + 2 NH3 + β-alanine + R1P
28	-DPYD1 -NM3 -UMPK1 -UMPK2 DPYS1 ECGF2 NT5C2 UPB1 CDA2 SHAPY1 irreversible	3 ADP + NADPH2 + 5 H2O + dCTP = 3 ATP + CO2 + Pi + NADP + 2 NH3 + β-alanine + dR1P
29	-DPYD2 -NM3 -DTYMK2 -UMPK1 -UMPK2 DPYS2 TYMS ECGF1 UPB2 DCTD SHAPY1 irreversible	4 ADP + NADPH2 + 4 H2O + ME4HF + dCTP = 4 ATP + CO2 + NADP + 2 NH3 + AisoB + DHF + dR1P
30	-DPYD2 -NM3 -UMPK1 -UMPK2 DPYS2 TYMS ECGF1 NT5C1 UPB2 DCTD SHAPY1 irreversible	3 ADP + NADPH2 + 5 H2O + ME4HF + dCTP = 3 ATP + CO2 + Pi + NADP + 2 NH3 + AisoB + DHF + dR1P
31	-DPYD2 -NM3 -DTYMK2 -UMPK1 -UMPK2 DPYS2 TYMS ECGF1 TK2 NT5C2 UPB2 CDA2 SHAPY1 irreversible	3 ADP + NADPH2 + 5 H2O + ME4HF + dCTP = 3 ATP + CO2 + Pi + NADP + 2 NH3 + AisoB + DHF + dR1P
32	-DPYD2 -NM3 -UMPK1 -UMPK2 DPYS2 TYMS ECGF1 TK2 NT5C1 NT5C2 UPB2 CDA2 SHAPY1 irreversible	2 ADP + NADPH2 + 6 H2O + ME4HF + dCTP = 2 ATP + CO2 + 2 Pi + NADP + 2 NH3 + AisoB + DHF + dR1P
33	-UMPK1 -UMPK2 UMPK3 2*SHAPY*1 irreversible	ADP + 2 H2O = ATP + 2 Pi
34	-DPYD1 UP -NM3 DTYMK3 -UMPK1 -UMPK2 DPYS1 TXNRD1 RRM1 NT5C4 UPB1 DCTD 2 *SHAPY*1 irre-versible	2 ADP + NADPH2 + 7 H2O + dCTP = 2 ATP + CO2 + 2 Pi + H + NADPH + 2 NH3 + β-alanine + R1P

No	Enzymes	overall reaction
35	-DPYD1 UP -NM3 DTYMK3 -UMPK1 -UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C2 NT5C4 UPB1 CDA2 2*SHAPY*1 irreversible	ADP + NADPH2 + 8 H2O + dCTP = ATP + CO2 + 3 Pi + H + NADPH + 2 NH3 + β-alanine + R1P
36	-DPYD1 NM1 -NM3 -UMPK1 -UMPK2 DPYS1 NP NT5C2 UPB1 CDA2 ITPA1 irreversible	2 ADP + NADPH2 + 5 H2O + dCTP = 2 ATP + PPi + CO2 + NADP + 2 NH3 + β-alanine + R1P
37	-DPYD1 NM1 -NM3 -UMPK1 -UMPK2 DPYS1 ECGF2 NT5C2 UPB1 CDA2 ITPA1 irreversible	2 ADP + NADPH2 + 5 H2O + dCTP = 2 ATP + PPi + CO2 + NADP + 2 NH3 + β-alanine + dR1P
38	-DPYD2 NM1 -NM3 -DTYMK2 -UMPK1 -UMPK2 DPYS2 TYMS ECGF1 UPB2 DCTD ITPA1 irreversible	3 ADP + Pi + NADPH2 + 4 H2O + ME4HF + dCTP = 3 ATP + PPi + CO2 + NADP + 2 NH3 + AisoB + DHF + dR1P
39	-DPYD2 NM1 -NM3 -UMPK1 -UMPK2 DPYS2 TYMS ECGF1 NT5C1 UPB2 DCTD ITPA1 irreversible	2 ADP + NADPH2 + 5 H2O + ME4HF + dCTP = 2 ATP + PPi + CO2 + NADP + 2 NH3 + AisoB + DHF + dR1P
40	-DPYD2 NM1 -NM3 -DTYMK2 -UMPK1 -UMPK2 DPYS2 TYMS ECGF1 TK2 NT5C2 UPB2 CDA2 ITPA1 irreversible	2 ADP + NADPH2 + 5 H2O + ME4HF + dCTP = 2 ATP + PPi + CO2 + NADP + 2 NH3 + AisoB + DHF + dR1P
41	-DPYD2 NM1 -NM3 -UMPK1 -UMPK2 DPYS2 TYMS ECGF1 TK2 NT5C1 NT5C2 UPB2 CDA2 ITPA1 irreversible	ADP + NADPH2 + 6 H2O + ME4HF + dCTP = ATP + PPi + CO2 + Pi + NADP + 2 NH3 + AisoB + DHF + dR1P
42	2*NM*1 -UMPK1 -UMPK2 UMPK3 2*ITPA*1 irreversible	ATP + 2 H2O = ADP + 2 PPi
43	-DPYD1 UP 2*NM*1 -NM3 DTYMK3 -UMPK1 -UMPK2 DPYS1 TXNRD1 RRM1 NT5C4 UPB1 DCTD (2 ITPA1) irreversible	NADPH2 + 7 H2O + dCTP = 2 PPi + CO2 + H + NADPH + 2 NH3 + β-alanine + R1P
44	-DPYD1 UP (2 NM1) -NM3 DTYMK3 -UMPK1 -UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C2 NT5C4 UPB1 CDA2 (2 ITPA1) irreversible	ATP + NADPH2 + 8 H2O + dCTP = ADP + 2 PPi + CO2 + Pi + H + NADPH + 2 NH3 + β-alanine + R1P

No	Enzymes	overall reaction
45	NM2 SHAPY4 irreversible	ATP + H2O = ADP + Pi
46	-DPYD1 DHODH UP -UMPS1 NM1 -NM2 -UMPK1 DPYS1 -CAD1 CAD3 UMPS2 NT5C3 UPB1 CDA1 CTPS CAD2 irreversible	PRPP + Gln + 2 ATP + NADPH2 + O2 + 4 H2O + aspartate = Glu + 2 ADP + PPi + CO2 + 3 Pi + NADP + H2O2 + NH3 + β-alanine + R1P
47	-DPYD1 DHODH UP -UMPS1 NM1 -UMPK1 DPYS1 - CAD1 CAD3 UMPS2 NT5C3 UPB1 CDA1 SHAPY4 CTPS CAD2 irreversible	PRPP + Gln + 3 ATP + NADPH2 + O2 + 5 H2O + aspartate = Glu + 3 ADP + PPi + CO2 + 4 Pi + NADP + H2O2 + NH3 + β-alanine + R1P
48	-DPYD1 UP -NM3 DPYS1 TXNRD1 RRM2 NT5C3 UPB1 CDA1 SHAPY2 irreversible	ADP + NADPH2 + 6 H2O + dCTP = ATP + CO2 + Pi + H + NADPH + 2 NH3 + β-alanine + R1P
49	-DPYD1 UP -NM3 -UMPK1 DPYS1 TXNRD1 RRM2 NT5C3 UPB1 CDA1 SHAPY1 irreversible	2 ADP + NADPH2 + 6 H2O + dCTP = 2 ATP + CO2 + Pi + H + NADPH + 2 NH3 + β-alanine + R1P
50	-DPYD1 UP NM1 -NM3 -UMPK1 DPYS1 TXNRD1 RRM2 NT5C3 UPB1 CDA1 ITPA1 irreversible	ADP + NADPH2 + 6 H2O + dCTP = ATP + PPi + CO2 + H + NADPH + 2 NH3 + β-alanine + R1P
51	-DPYD1 UP NM1 -NM2 -NM3 DTYMK3 -UMPK2 DPYS1 TXNRD1 RRM1 NT5C3 UPB1 CDA1 DCTD (2 SHAPY2) CTPS irreversible	NADPH2 + 8 H2O + dCTP = CO2 + 3 Pi + H + NADPH + 2 NH3 + β-alanine + R1P
52	-DPYD1 UP NM1 -NM3 DTYMK3 -UMPK2 DPYS1 TXNRD1 RRM1 NT5C3 UPB1 CDA1 DCTD (2 SHAPY2) SHAPY4 CTPS irreversible	ATP + NADPH2 + 9 H2O + dCTP = ADP + CO2 + 4 Pi + H + NADPH + 2 NH3 + β-alanine + R1P
53	-DPYD1 UP NM1 -NM2 -NM3 DTYMK3 -UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C2 NT5C3 UPB1 CDA1 CDA2 (2 SHAPY2) CTPS irreversible	ATP + NADPH2 + 9 H2O + dCTP = ADP + CO2 + 4 Pi + H + NADPH + 2 NH3 + β-alanine + R1P
54	-DPYD1 UP NM1 -NM3 DTYMK3 -UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C2 NT5C3 UPB1 CDA1 CDA2 (2 SHAPY2) SHAPY4 CTPS irreversible	2 ATP + NADPH2 + 10 H2O + dCTP = 2 ADP + CO2 + 5 Pi + H + NADPH + 2 NH3 + β-alanine + R1P

No	Enzymes	overall reaction
55	-DPYD1 DHODH UP -UMPS1 NM1 -NM2 -UMPK2 DPYS1 -CAD1 CAD3 UMPS2 UMPK3 NT5C3 UPB1 CDA1 (2 SHAPY2) CTPS CAD2 irreversible	PRPP + Gln + 3 ATP + NADPH2 + O2 + 6 H_2O + aspartate = Glu + 3 ADP + PPi + CO_2 + 5 Pi + NADP + H_2O_2 + NH_3 + β-alanine + R1P
56	-DPYD1 DHODH UP -UMPS1 NM1 -UMPK2 DPYS1 -CAD1 CAD3 UMPS2 UMPK3 NT5C3 UPB1 CDA1 (2 SHAPY2) SHAPY4 CTPS CAD2 irreversible	PRPP + Gln + 4 ATP + NADPH2 + O2 + 7 H_2O + aspartate = Glu + 4 ADP + PPi + CO_2 + 6 Pi + NADP + H_2O_2 + NH_3 + β-alanine + R1P
57	-DPYD1 UP NM1 -NM2 -NM3 DTYMK3 (-2 UMPK1) -UMPK2 DPYS1 TXNRD1 RRM1 NT5C3 UPB1 CDA1 DCTD (2 SHAPY1) CTPS irreversible	2 ADP + NADPH2 + 8 H_2O + dCTP = 2 ATP + CO_2 + 3 Pi + H + NADPH + 2 NH_3 + β-alanine + R1P
58	-DPYD1 UP NM1 -NM3 DTYMK3 (-2 UMPK1) -UMPK2 DPYS1 TXNRD1 RRM1 NT5C3 UPB1 CDA1 DCTD (2 SHAPY1) SHAPY4 CTPS irreversible	ADP + NADPH2 + 9 H_2O + dCTP = ATP + CO_2 + 4 Pi + H + NADPH + 2 NH_3 + β-alanine + R1P
59	-DPYD1 UP NM1 -NM2 -NM3 DTYMK3 (-2 UMPK1) -UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C2 NT5C3 UPB1 CDA1 CDA2 (2 SHAPY1) CTPS irreversible	ADP + NADPH2 + 9 H_2O + dCTP = ATP + CO_2 + 4 Pi + H + NADPH + 2 NH_3 + β-alanine + R1P
60	-DPYD1 UP NM1 -NM3 DTYMK3 (-2 UMPK1) -UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C3 UPB1 CDA1 CDA2 (2 SHAPY1) SHAPY4 CTPS irreversible	NADPH2 + 10 H_2O + dCTP = CO_2 + 5 Pi + H + NADPH + 2 NH_3 + β-alanine + R1P
61	-DPYD1 UP (3 NM1) -NM2 -NM3 DTYMK3 (-2 UMPK1) -UMPK2 DPYS1 TXNRD1 RRM1 NT5C3 UPB1 CDA1 DCTD (2 ITPA1) CTPS irreversible	NADPH2 + 8 H_2O + dCTP = 2 PPi + CO_2 + Pi + H + NADPH + 2 NH_3 + β-alanine + R1P
62	-DPYD1 UP (3 NM1) -NM3 DTYMK3 (-2 UMPK1) -UMPK2 DPYS1 TXNRD1 RRM1 NT5C3 UPB1 CDA1 DCTD SHAPY4 (2 ITPA1) CTPS irreversible	ATP + NADPH2 + 9 H_2O + dCTP = ADP + 2 PPi + CO_2 + 2 Pi + H + NADPH + 2 NH_3 + β-alanine + R1P

No	Enzymes	overall reaction
63	-DPYD1 UP 3NM1 -NM2 -NM3 DTYMK3 $-2UMPK1$ -UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C2 NT5C3 UPB1 CDA1 CDA2 2$ITPA$1 CTPS irreversible	ATP + NADPH2 + 9 H_2O + dCTP = ADP + 2 PPi + CO_2 + 2 Pi + H + NADPH + 2 NH3 + β-alanine + R1P
64	-DPYD1 UP (3 NM1) -NM3 DTYMK3 (-2 UMPK1) - UMPK2 DPYS1 TXNRD1 RRM1 TK2 NT5C2 NT5C3 UPB1 CDA1 CDA2 SHAPY4 (2 ITPA1) CTPS irreversible	2 ATP + NADPH2 + 10 H_2O + dCTP = 2 ADP + 2 PPi + CO_2 + 3 Pi + H + NADPH + 2 NH3 + β-alanine + R1P

No	Enzyme Subsets*	Overall reaction of enzyme subsets
1	-DPYD1 DPYS1 UPB1 irreversible	uracil + NADPH2 + 2 H2O = CO2 + NADP + NH3 + β-alanine
2	-DPYD2 DPYS2 TYMS ECGF1 UPB2 irreversible	thymidine + dUMP + Pi + NADPH2 + 2 H2O + ME4HF = dTMP + CO2 + NADP + NH3 + AisoB + DHF + dR1P
3	DHODH -UMPS1 -CAD1 CAD3 UMPS2 CAD2 irreversible	PRPP + Gln + 2 ATP + O2 + aspartate = UMP + Glu + 2 ADP + PPi + 2 Pi + H2O2
4	UP reversible	uridine + Pi = uracil + R1P
5	NM5 SHAPY5 irreversible	ATP + H2O = ADP + Pi
6	DTYMK1 SHAPY6 irreversible	ATP + H2O = ADP + Pi
7	NT5C3 CDA1 irreversible	CMP + 2 H2O = uridine + Pi + NH3 *All other enzyme subsets are formed by one enzyme each.

Appendix G

Procedure for extracting the tree having as root a given node from the given graph written in pseudo-code.

$Procedure BildTree(root, direction)$
begin
  $CurrentNode = root;$
  $UnexploredNodesList = CreateListUnexploredNodes(CurrentNode, direction);$
  while $(UnexploredNodesList not empty)$
      begin
        $Node = ExtractNode(UnexploredNodesList);$
       if $(CheckCycle(CurrentNode, Node, Tree, direction))$
         begin
          if $(NewCycle(CurrentNode, Node, CycleList))$
            begin   **end**;
        else
            begin $AddToCycleList(CurrentNode, Node, CycleList);$ **end**;
         end
      else
         begin
          if $(OldArrow(CurrentNode, Node, Tree, direction))$
            begin   **end**;
         else
            begin
             if $(FirstArrow(CurrentNode, Node, Tree) == true)$
              begin $AddToSon(CurrentNode, Node);$ **end**;
            else
              begin $AddToBrother(CurrentNode, Node);$ **end**;
            end;
          end;
        fi
       end;
      end.

Appendix H

Backtracking procedure (in pseudocode), that builds and classifies route after route having as start point a given factor or target.

```
ProcedureBt(StartPlace)
begin Step := 1;
   InitialiseLists();
   StoreCurentState(Step);
   if (CheckConflict(StartPlace) == 1)SolveConflict();
else begin
        CurrentPlace[Step] := StartPlace; Direction := 1;
        SetSolutionSets(Step, SolutionSet1, SolutionSet2, Direction, TypeT);
     end;
     while (Step > 0)
           while ((SolutionSet1notempty)or(SolutionSet2notempty))
                 begin
                  if (ChangedRoute[Step] == 1)RestoreState();
                     StoreCurentState(); ChoseTransition(Step, Direction, TypeT);
                     switch(Solution(Step))
                     begin
                       case1 :
                       begin if NewSolution()
                                begin
                                   AddToRoute(); StoreAdditionalInformation();
                                end;
                       end;
                       case − 1 :
                       begin
                          if NewSolution()
                             begin
                             AddToIncompleteRoute(); StoreAdditionalInformation();
                             end;
                       end;
                       case2 :
                       begin
                          if NewSolution()
                             begin
                                AddToCycle(); StoreAdditionalInformation();
                             end;
                       end;
```

```
case − 2 :
begin
  if NewSolution()
    begin
      AddToIncompleteCycle();  StoreAdditionalInformation();
    end;
  end;
  case0 :
  begin
    CurrentPlace[Step + 1] = ExtractPlaceToBeExpand(Step, direction, TypeT);
    if (CurrentPlace[Step + 1]! = −1)
      begin
        if (CheckConflict(CurrentPlace[Step + 1]) == 1)SolveConflict();
      else
          begin
            Step + +;  ChangedRoute[Step] = 0;
            SetSolutionSets(Step, SolutionSet1, SolutionSet2, Direction, TypeT);
          end;
        end;
      end;
    end;
  end;
  if (ChangedRoute[Step] == 1)BackRoute();
    Step − −;
    if (ChangedRoute[Step] == 1)BackRoute();
      ChangedRoute[Step] == 0;  RestoreState();
    end;
```

Bibliography

[AD98]   H. Alla and R. David. Continuous and hybrid petri nets. *J. Circ. Syst. Comp.*, 8:159–188, 1998. 53

[AJ83]   A.V. Aho and D.U. Jeffrey. *Data Structures and Algorithms*. Addison-Wesley, Reading, Massachusetts, 1983. 79

[Alb94]   R.A. Alberty. Constraints in biochemical reactions. *Biophys. Chem.*, 49:251–261, 1994. 18

[BBK⁺76]   S. Brosh, P. Boer, B. Kupfer, A. deVries, and O. Sperling. *De novo* synthesis of purine nucleotides in human peripheral blood leukocytes. excessive activity of the pathway in hypoxanthine-guanine phosphoribosyltransferase deficiency. *J. Clin. Invest.*, 58:289–297, 1976. 38

[BDDL⁺98] P. Bork, T. Dandekar, Y Diaz-Lazcoz, F. Eisenhaber, M. Huynen, and Y. Yuan. Predicting function: From genes to genomes and back. *J. Mol. Biol*, 283:707–725, 1998. 4, 106

[Bec00]   M.A. Becker. Hyperuricemia and gout. In C.R. Scriver, W.S. Sly, B. Childs, A.L. Beaudet, D. Valle, K.W. Kinzler, and B. Vogelstein, editors, *The Metabolic and Molecular Bases of Inherited Disease*, volume 2, pages 2513–2532. McGraw-Hill, New York, 2000. 42

[BHB90]   G.C. Brown, R.P. Hafner, and M.D. Brand. A 'top-down' approach to determination of control coefficients in metabolic control theory. *Eur. J. Biochem*, 188:321–325, 1990. 1, 2

[BI99]   U.S. Bhalla and R. Iyengar. Emergent properties of networks of biological signalling pathways. *Science*, 28:381–387, 1999. 4, 10

[Bor00]   L.G. Boros. Cancer rates and population thimin status. *Anticancer Res.*, 20:2245–48, 2000. 49

[BSM⁺00]   I.A. Bykova, O.N. Solovjeva, L.E. Meshalkina, M.V. Kovina, and G. Kochetov. One-substrate transketolase - catalysed reaction. *Biochem. Biophys. Res. Comm.*, 280:845–847, 2000. 101

[BTS02]   J. Berg, J. Tymoczko, and L. Stryer. *Biochemistry*. Freeman, New York, 5 edition, 2002. 11

[BWtK⁺99]   B. Bakker, M.C. Walsh, B.H. ter Kuile, F.I.C. Mensonides, P.A.M. Michels, F.R. Opperdoes, and Westerhoff H.V. Contributions of glucose transport to the control of the glycolytic flux in *Trypanosoma brucei*. In *Proc. Natl Acad. Sci. USA*, volume 96, pages 10098–10103, 1999. 68

[Cam99]   K.S. Campbell. Signal transduction from the b cell antigen-receptor. *Curr. Opin. Immunol.*, 11:256–264, 1999. 12, 96, 108

[CBHC95]   A. Cornish-Bowden, J.-H.S. Hofmeyr, and M.L. Cárdenas. Strategies for manipulating metabolic fluxes in biotechnology. *Bioorg. Chem.*, 23:439–449, 1995. 88

[CCV72]   M.G. Colomb, A. Cheruy, and P.V. Vignais. Nucleoside diphosphokinase from beef heart cytosol. ii. characterization of the phosphorylated intermediate. *Biochemistry*, 11:3378–3386, 1972. 30

[CCV⁺00]   M. Cascante, J.J. Centelles, R.L. Veech, W-n.P. Lee, and L.G. Boros. Role of thiamin (vitamin b1) and transketolase in tumor cell proliferation. *Nutr. Canc.*, 36:150–154, 2000. 17, 49, 103

[CJ01]   I. Coppens and K.A. Joiner. *Parasite-host cell interactions in toxoplasmosis: new avenues for intervertions?* Cambridge University Press, 2001. 51

[Cla80]   B.L. Clarke. Stability of complex reaction networks. *Adv. Chem. Phys.*, 43:1–216, 1980. 9, 61, 63

[CLR90]   T.H. Cormen, C.E. Leiserson, and Rivest R.L. *Introduction to Algorithms*. McGraw-Hill, New York, 1990. 79

[CLW02]   R.I. Christopherson, S.D. Lyons, and P.K. Wilson. Inhibitors of the novo nucleotide biosynthesis as drugs. *Acc. Chem. Res*, 35(11):961–971, 2002. 8, 48

[Com72]   F. Commoner. Deadlocks in petri nets. Applied data research, Inc. Wakefield, Massachusetts, 1972. 67

[CS90]   J.M. Colom and M. Silva. Convex geometry and semiflows in p/t nets. a comparative study of algorithms for computation of minimal p-semiflows. In Rozenberg G., editor, *Advances in Petri Nets*, volume 2, pages 79–112. Springer, Berlin, 1990. 62, 64

[CX97]   F. Chu and X.-L. Xie. Deadlock analysis of petri nets using siphons and mathematical programming. In *IEEE Trans. on Robotics and Automation*, volume 13, pages 793–804, 1997. 66

[DC98]   J. D'Ari and J. Casadesus. Underground metabolism. *Bioessays*, 20:181–186, 1998. 99

[DeF95]   A.L. DeFranco. Transmembrane signalling by antigen receptors of b and t lymphocyte. *Current Opinion in Cell Biology*, 7:163–175, 1995. 96

[DeF97]   A.L. DeFranco. The complexity of signalling pathways activated by the bcr. *Current Opinion in Immunology*, 9:296–308, 1997. 96

[DlFSWM02] A. De la Fuente, J.L. Snoep, H.V. Westerhoff, and P. Mendes. Metabolic control in integrated biochemical systems. *Eur. J. Biochem.*, 269:4399–4408, 2002. 3

[DMS99]   P.V. Danenberg, H. Malli, and S. Swenson. Thymidylate synthase inhibitors. *Semin. Oncol.*, 26(6):621–631, 1999. 48

[DSS+99]   T. Dandekar, S. Schuster, B. Snel, M. Huynen, and P. Bork. Pathway alignment: Application to the comparative analysis of glycolytic enzymes. *Biochem. J.*, 343:115–124, 1999. 4, 8, 106

[ET89]   P. Érdi and J. Tóth. *Mathematical Models of Chemical Reactions.* Manchester University Press, Manchester, 1989. 62

[FB02]   B.A. Fox and D.J. Bzik. *De novo* pyrimidine biosynthesis is required for virulence of *Toxoplasma gondii*. *Nature*, 415:926–929, 2002. 8, 52

[Fel92]   D.A. Fell. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem. J.*, 286:313–330, 1992. 1, 2

[FG99]   P. Franchetti and M Grifantini. Nucleoside and non-nucleoside imp dehydrogenase inhibitors as antitumor and antiviral agents. *Curr. Med. Chem.*, 6(7):599–614, 1999. 8, 48

[FGN02]   J. Förster, A.K. Gombert, and J. Nielsen. A functional genomics approach using metabolomics and in silico pathway analysis. *Biotechnol. Bioeng.*, 79:703–12, 2002. 106

[FGS+01]   E. Fiedler, R. Golbik, G. Schneider, K. Tittmann, H. Neef, S. Konig, and G. Hubner. Examination of donor substrate conversion in yeast transketolase. *J. Biol. Chem.*, 276:16051–16058, 2001. 23, 101

[FHF+99]   R.I. Fox, M.L. Herrmann, C.G. Frangou, G.M. Wahl, R.E. Morris, V. Strand, and B.J Kirschbaum. Mechanism of action for leflunomide in rheumatoid arthritis. *Clin. Immunol.*, 93:198–208, 1999. 46

[FS86]   D.A. Fell and J.R. Small. Fat synthesis in adipose tissue. an ex-
            amination of stoichiometric constrains. *Biochem. J.*, 238:781–786,
            1986. 2

[FW00]   D.A. Fell and A. Wagner. The small world of metabolism. *Nat.
            Biotechnol.*, 18:1121–1122, 2000. 2, 3, 10

[GDB90]   A. Goldbeter, G. Dupont, and M.J. Berridge. Minimal model for
            signal-induced ca^{2+} oscillations and for their frequency encoding
            through protein phosphorylation. In *Proc. Natl. Acad. Sci. USA*,
            volume 87, pages 1461–1465, 1990. 2

[GJCM95]   W.-Y. Gao, D.G. Johns, S. Chokekijchai, and H. Mitsuya. Dis-
            parate actions of hydroxyurea in potentiation of purine and pyrim-
            idine 2,3-dideoxynucleoside activities against replication of human
            immunodeficiency virus. In *Proc. Natl. Acad. Sci. USA*, volume 92,
            pages 8333–7, 1995. 8, 50

[GKV01]   H. Genrich, R. Küffner, and K. Voss. Executable petri net models
            for the analysis of metabolic pathways. *Int. J. STTT*, 3:394–404,
            2001. 10, 53

[Gol96]   A. Goldbeter. *Biochemical Oscillations and Cellular Rhythms.*
            Cambridge University Press, Cambridge, 1996. 1, 12, 81, 88

[Gol02]   A. Goldbeter. Computational approaches to cellurar rhythms. *Na-
            ture*, 420:238–245, 2002. 2

[HEB⁺01]   S. Helfert, A.M. Estevez, B. Bakker, P. Michels, and C. Clayton.
            Roles of triosephosphate isomerase and aerobic metabolism in *Try-
            panosoma brucei. Biochem. J.*, 357:117–125, 2001. 68

[HHDG85]   H.-G. Holzhütter, W. Henke, W. Dubiel, and G. Gerber. A mathe-
            matical model to study short-term regulation of mitochondrial en-
            ergy transduction. *Biochim. Biophys. Acta*, 810:252–268, 1985. 1,
            2

[HJ72]   F. Horn and R. Jackson. General mass action kinetics. *Arch. Ra-
            tional Mech. Anal.*, 47:81–116, 1972. 9, 62

[HJXC01]   Y. Huang, M.D. Jeng, Z. Xie, and S. Chung. Deadlock prevention
            policy based on petri nets and siphons. *Arch. Rational Mech. Anal.*,
            39(2):283–305, 2001. 66

[HKS00]   M. Heiner, I. Koch, and S. Schuster. Using time-dependent petri
            nets for the analysis of metabolic networks. In R. Hofestädt,

K. Lautenbach, and M. Lange, editors, *Modellierung und Simulation Metabolischer Netzwerke*, volume 10, pages 15–21. McGraw-Hill, Faculty of Computer Science, University of Magdeburg, 2000. 10, 59

[HKV01]   M. Heiner, I. Koch, and K. Voss. Analysis and simulation of steady states in metabolic pathways with petri nets. In Jensen K, editor, *CPN '01 - Third Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools*, pages 15–34, Denmark, 2001. University of Aarhus. 10

[HM00]   M.S. Hershfield and B.S. Mitchell. Inummodeficiency disease caused by adenosine deaminase deficiency and purine nucleoside phosphorylase deficiency. In C.R. Scriver, W.S. Sly, B. Childs, A.L. Beaudet, D. Valle, K.W. Kinzler, and B. Vogelstein, editors, *The Metabolic and Molecular Bases of Inherited Disease*, volume 2, pages 2585–2612. McGraw-Hill, New York, 2000. 8, 44

[HNR02]   R. Heinrich, B.G. Neel, and T.A. Rapoport. Mathematical models of protein kinase signal transduction. *Mol. Cell.*, 9:957–970, 2002. 4, 10, 11, 75

[Hof94]   R. Hofestädt. A petri net application to model metabolic processes. *Syst. Anal. Mod. Simul.*, 16:113–122, 1994. 10, 82

[HR74]   R. Heinrich and T.A. Rapoport. A linear steady-state treatment of enzymatic chains. general properties, control and effector strength. *Eur. J. Biochem.*, 42:89–95, 1974. 1, 2

[HS96]   R. Heinrich and S. Schuster. *The Regulation of Cellular Systems*. Chapman and Hall, New York, 1996. 1, 5, 10, 18, 56, 58, 61, 63

[HSG$^+$93]   H.G. Holzhütter, A. Schwendel, T. Grune, J. Quedenau, and W. Siems. Estimation of steady-state flux rates in metabolic systems by computer simulations of radioactive tracer experiments. *Comput Appl Biosci*, 9(5):573–80, 1993. 35

[HSH91]   R. Heinrich, S. Schuster, and H.-G. Holzhütter. Mathematical analysis of enzymic reaction systems using optimization principles. *Eur. J. Biochem.*, 201:1–21, 1991. 2, 3

[IMA00]   M.V. Iordache, J.O. Moody, and P.J. Antsaklis. Automated synthesis of liveness enforcing supervisors using petri nets. Technical report, Dept. of Electrical Engr., Univ. of Notre Dame, 2000. 66

[Jen97]   K. Jensen. *Coloured Petri nets. Basic concepts, analysis, methods and practical use. Monographs in theoretical computer science*, volume 1-3. Springer-Verlag, Berlin, Heidelberg, New York, 1992-1997. 54

[Jen98]   K. Jensen. A brief introduction to coloured petri nets. In *Proc. Workshop on the Applicability of Formal Models*, pages 55–58. Aarhus, Denmark, 1998. 14

[JF00]   H.A. Jinnah and T. Friedmann. Lesch-nyhan disease and its variants. In C.R. Scriver, W.S. Sly, B. Childs, A.L. Beaudet, D. Valle, K.W. Kinzler, and B. Vogelstein, editors, *The Metabolic and Molecular Bases of Inherited Disease*, volume 2, pages 2537–2560. McGraw-Hill, New York, 2000. 9, 43

[JP89]   A. Joshi and B.O. Palsson. Metabolic dynamics in the human red cell. part i–a comprehensive kinetic model. *J. Theor. Biol.*, 141:515–28, 1989. 35, 38

[JTA⁺00]   H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000. 2, 3, 4, 10, 13, 40

[Kar77]   P. Karlson. *Kurzes Lehrbuch der Biochemie für Mediziner und Naturwissenschaftler*. Georg Thieme Verlag, Stuttgart, 1977. 4

[KB73]   H Kacser and J.A. Burns. The control of flux. Department of Genetics, University of Edinburg, 1973. 1, 2

[Kem93]   P. Kemper. Linear time algorithm to find a minimal deadlock in a strongly connected free-choice net. In M. Ajmone-Marsan, editor, *Proc. 14th International Conference Application and Theory of Petri Nets*, volume 691 of *LNCS*, pages 319– 338, Chicago, 1993. Springer. 66

[KG91]   M. Kanehisa and S. Goto. Kegg: Kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.*, 28:27–30, 1991. 104, 106

[KHWB97]   B.N. Kholodenko, J.B. Hoek, H.V. Westerhoff, and G.C. Brown. Quantification of information transfer via cellular signal transduction pathways. *FEBS Lett.*, 414:30–434, 1997. 2, 11, 75

[KKB⁺02]   B.N. Kholodenko, A. Kiyatkin, F.J. Bruggeman, E. Sontag, H.V. Westerhoff, and J.B. Hoek. Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. In *Proc. Natl. Acad. Sci. USA*, volume 99, pages 12841–12846, 2002. 4, 10

[Knu74]   D. Knuth. *The Art of Computer Programming*, volume Sorting and Searching. Technical publishing house, 1974. 79

[Kor95]   B.E. Korba. In vitro evaluation of combination therapies against hepatitis b virus replication. *Antiviral Research*, 29:49–51, 1995. 50

[KS98]   S. Klamt and J. Stelling. Two approaches for metabolic pathway analysis? *Trends Biotechn.*, 21:64–69, 1998. 105

[KSYA96]   S. Koyano, T. Suzutani, I. Yoshida, and M. Azuma. Analysis of phosphorylation pathways of antiherpesvirus nucleosides vy varicella-zoster virus-specific enzymes. *Antimicrobial Agents and Chemotherapy*, 40(4):920–923, 1996. 51

[KSYA00]   S. Koyano, T. Suzutani, I. Yoshida, and M. Azuma. Pathway analysis in metabolic databases via differential metabolic display (dmd). *Bioinformatics*, 16:825–836, 2000. 10

[KW91]   D. Kahn and H.V. Westerhoff. Control theory of regulatory cascades. *J. Theor. Biol.*, 153(2):255–85, 1991. 3

[LA75]   A.S. Liacouras and E.P. Anderson. Uridine-cytidine kinase. *Arch. Biochem. Biophys.*, 168:66–73, 1975. 6, 20

[LB87]   J. Leiser and J.J. Blum. On the analysis of substrate cycles in large metabolic systems. *Cell Biophys.*, 11:123–138, 1987. 4, 10

[LG86]   L. Livovschi and H. Georgescu. *Algorithms' synthesis and analysis*. Scientific and enciclopedic publishing house, Bucharest, 1986. 90

[LGMA75]   A.S. Liacouras, T.Q. Garvey, F.K. Millar, and E.P Anderson. Uridine-cytidine kinase. kinetic studies and reaction mechanism. *Arch. Biochem. Biophys.*, 168:74–80, 1975. 6, 10, 20

[LHC96]   J.C Liao, S.Y. Hou, and Y.P. Chao. Pathway analysis, engineering and physiological considerations for redirecting central metabolism. *Biotechnol. Bioeng.*, 52:129–140, 1996. 4, 5

[MA98]   J.O. Moody and P.J. Antsaklis. Deadlock avoidance in petri nets with uncontrollable transitions. In *Proceedings of 1998 American Control Conference*, pages 24–26, Philadelphia, PA, 1998. 66

[MCS98]   J.E. Mittenthal, A.Y.B. Clarke, and A. Scheeline. Designing metabolism: Alternative connectivities for the pentose phosphate pathway. *Bul. Of Mat. Biol.*, 60:815–856, 1998. 2

[MDNM00]   H. Matsuno, A. Doi, M. Nagasaki, and S. Miyano. Hybrid petri net representation of gene regulatory network. *Pacif. Symp. Biocomp.*, 5:341–352, 2000. 53

[Men97]   P. Mendes. Biochemistry by numbers: simulation of biochemical pathways with gepasi 3. *Trends Biochem. Sci.*, 22:361–363, 1997. 1, 66

[MH99]   E. Melendez-Hevia. The game of the pentose phosphate cycle: A mathematical approach to study the optimization in design of metabolic pathways during evolution. *Biomed. Biochim. Acta*, 49:903–916, 1999. 2, 3, 17

[MHTSK90]   E. Melendez-Hevia, N.V. Torres, J. Sicilia, and H. Kaiser. Control analysis of transition times in metabolic systems. *Biochem. J.*, 265:195–202, 1990. 1, 2

[MKK01]   P. Mazzotta, A. Kwasnicka, and G.J. Kutas. Cancer chemotherapy: The role of pharmacological agents in the management of haematological malignancies. *Pacif. Symp. Biocomp.*, 79:38–45, 2001. 8, 48

[MMHC99]   R. Melendez, E. Melendez-Hevia, and E.I. Canela. The fractal stucture of glycogen: A clever solution to optimise cell metabolism. *Biochem. J.*, 77:1327–1332, 1999. 2, 3

[MS88]   T. Meyer and L. Stryer. Molecular model for receptor-stimulating calcium spiking. In *Proc. Natl. Acad. Sci. USA*, volume 85, pages 5051–5055, 1988. 1, 2

[MWY$^+$03]   Y. Mizutani, H. Wada, O. Yoshida, M. Fukushima, M. Nakao, and T. Miki. Significance of the thymidine kinase activity in the renal cell carcinoma. *J. Urol.*, 169(2):706–9, 2003. 47

[NCO$^+$97]   L. Nitschke, R. Carsetti, B. Ocker, G. Köhler, and M. Lamers. Cd22 is a negative regulator of b-cell receptor signalling. *Current Biology*, 7:133–143, 1997. 98

[NnSVPI$^+$97] J.C. Nuño, I. Sánchez-Valdenebro, C. Pérez-Iratxeta, E. Meléndez-Hevia, and F. Montero. Network organization of cell metabolism: monosaccharide interconversion. *Biochem. J.*, 324:103–111, 1997. 5, 6, 21, 23, 100, 101

[OBJO&D01]   J.S. Oliveira, C.G. Bailey, J.B. Jones-Oliveira, and D.A. & Dixon. An algebraic-combinatorial model for the identification and mapping of biochemical pathways. *Bull. Math. Biol.*, 63:163–196, 2001. 2, 10, 82

[OBK$^+$02]   K.M. Overkamp, B.M. Bakker, P. Kötter, M.A.H. Luttik, J.P. van Dijken, and J.T. Pronk. Metabolic engineering of glycerol production in *Saccharomyces cerevisiae*. *Applied and environmental microbiology*, 68(6):2814–2821, 2002. 68

[Ore68]   A. Orengo. Regulation of enzymic activity by metabolites. i. uridine-cytidine kinase of *Novikoff ascites* rat tumor. *J. Biol. Chem.*, 8:2204–2209, 1968. 20

[Oth81]   H.G. Othmer. The interaction of structure and dynamics in chemical reaction networks. In Deuflhard P Ebert KH and Jäger W, editors, *Modelling of Chemical Reaction Systems*, pages 2–19. Springer, Berlin, 1981. 67

[Pal00]   B.O. Palsson. The challenges of in silico biology. *Nature Biotechn.*, 18:1147–1150, 2000. 4, 10

[PCT85]   R.C. Payne, N. Cheng, and T.W. Traut. Uridine kinase from *Ehrlich Ascites Carcinoma*. purification and properties of homogeneous enzyme. *J. Biol. Chem.*, 18:10242–10247, 1985. 20

[PE02]   M.J. Pugmire and S.E. Earlick. Structural analysis reveal two distinct families of nucleoside phosphorylases. *Biochem. J.*, 361:1–25, 2002. 49

[PL85]   P.W. Postma and J.W. Lengeler. Phosphoenolpyruvate: carbohydrate phosphotransferase system of bacteria. *Microbiol. Rev.*, 49(3):232–69, 1985. 11

[PSVNn+99] T Pfeiffer, I. Sánchez-Valdenebro, J.C. Nuño, F. Montero, and S. Schuster. Metatool: For studying metabolic networks. *Bioinformatics*, 15:251–257, 1999. 2, 5, 20, 64, 67, 101, 106

[RB01]   J.M. Rohwer and F.C. Botha. Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data. *Biochem. J.*, 358:437–445, 2001. 4, 5, 64, 104

[Red88]   C. Reder. Metabolic control theory: a structural approach. *J. Theor. Biol.*, 135:175–201, 1988. 1, 2

[Rei85]   W. Reisig. *Petri Nets: An Introduction*, volume 3. Springer, Berlin, 1985. 9, 54, 61, 65, 67, 81

[RKL+87]   J.A. Royds, H.J. Kennedy, P.V. Little, C.B. Taylor, and D.R. Triger. Serum aldolase isoenzymes in benign and malignant liver disease. *Clin. Chim .Acta.*, 167:237–246, 1987. 18

[RLM96]   V.N. Reddy, M.N. Liebmann, and M.L. Mavrovouniotis. Qualitative analysis of biochemical reaction systems. *Comp. Biol. Med.*, 26:9–24, 1996. 82

[RSL00]   K.O. Raivio, M. Saksela, and R. Lapatto. Xanthine oxidoreductase
          – role in human pathophysiology and in hereditary xanthinuria. In
          C.R. Scriver, W.S. Sly, B. Childs, A.L. Beaudet, D. Valle, K.W.
          Kinzler, and B. Vogelstein, editors, *The Metabolic and Molecular
          Bases of Inherited Disease*, volume 2, pages 2639–2652. McGraw-
          Hill, New York, 2000. 43

[SAN98]   G.N. Stephanopoulos, A.A. Aristidou, and J. Nielsen. *Metabolic
          Engineering. Principles and Methodologies.* Academic Press, San
          Diego, 1998. 4

[SB88]   A. Seressiotis and J.E. Bailey. Mps: An artificially intelligent soft-
          ware system for the analysis and synthesis of metabolic pathways.
          Biotechnol. Bioeng., 31:587–602, 1988. 5, 100

[SBG$^+$96] E. Selkov, S. Basmanova, T. Gaasterland, I. Goryanin,
          Y. Gretchkin, N. Maltsev, V. Nenashev, R. Overbeek, E. Panyushk-
          ina, Jr.E. Pronevitch, L. Selkov, and I. Yunus. The metabolic path-
          way collection from emp: The enzymes and metabolic pathways
          database. *Nucl. Acids Res.*, 24:26–28, 1996. 106

[SBS91]   J.A. Sokoloski, G.P. Beardsley, and A.C. Sartorelli. Induction of hl-
          60 leukemia cell differentiation by tetrahydrofolate inhibitors of the
          novo purine nucleotide biosynthesis. *Cancer Chemother Pharmacol*,
          28:39–44, 1991. 8, 49

[Sch96]   K. Schmidt. Siphons and traps for algebraic petri nets. In *Proc.
          Workshop CS&P*, pages 157–168, Berlin, 1996. 65

[SDF99]   S. Schuster, T. Dandekar, and D.A. Fell. Detection of elementary
          flux modes in biochemical networks: a promising tool for pathway
          analysis and metabolic engineering. *Trends Biotechnol.*, 17:53–60,
          1999. 2, 4, 5, 64

[SFD00]   S. Schuster, D.A. Fell, and T. Dandekar. A general definition of
          metabolic pathways useful for systematic organization and analysis
          of complex metabolic networks. *Nature Biotechnol.*, 18:326–332,
          2000. 4, 5, 6, 11, 12, 25, 28, 64, 104, 105

[SGHS97] A. Schwendel, T. Grune, H.G. Holzhütter, and W.G. Siems. Models
          for the regulation of purine metabolism in rat hepatocytes: evalua-
          tion of tracer kinetic experiments. *Am J Physiol.*, 273:239–46, 1997.
          35

[SH91]   S. Schuster and T. Höfer. Determining all extreme semi-positive
          conservation relations in chemical reaction systems. a test criterion

for conservativity. *J. Chem. Soc. Faraday Trans.*, 87:2561–2566, 1991. 61, 62

[SH94] S. Schuster and C. Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.*, 2:165–182, 1994. 5, 12, 64, 101

[SH95a] R. Schuster and H.-G. Holzhütter. Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations. application to enzyme deficiencies of red blood cells. *Eur. J. Biochem.*, 229:403–418, 1995. 9

[SH95b] S. Schuster and C. Hilgetag. What information about the conserved-moiety structure of chemical reaction systems can be derived from their stoichiometry? *J. Phys. Chem.*, 99:8017–8023, 1995. 61, 62

[SHL⁺83] E.L. Smith, R.L. Hill, I.R. Lehman, R.J. Lefkowitz, P. Handler, and A. White. *Principles of Biochemistry. General Aspects.* McGraw-Hill, New York, 1983. 21

[SHWF02] S. Schuster, C. Hilgetag, J.H. Woods, and D.A. Fell. Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.*, 45:153–181, 2002. 5, 6, 10, 25, 64, 101

[SJH89] R. Schuster, G. Jacobasch, and H.-G. Holzhütter. Mathematical modelling of metabolic pathways affected by an enzyme deficiency. energy and redox metabolism of glucose-6-phosphate-dehydrogenase-deficient erythrocytes. *Eur. J. Biochem.*, 182:605–612, 1989. 9

[SKB⁺02] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E.D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 429:190–193, 2002. 6, 13

[SKW00] S. Schuster, B.N. Kholodenko, and H.V. Westerhoff. Cellular information transfer regarded from a stoichiometry and control analysis perspective. *Biosystems*, 55:73–81, 2000. 1, 4

[SKW⁺02] S. Schuster, S. Klamt, F. Weckwerth, F. Moldenhauer, and T. Pfeiffer. Use of network analysis of metabolic systems in bioengineering. *Biosystems*, 24(6):363–372, 2002. 10, 40

[SLP00] C.H. Schilling, D. Letscher, and B.O. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. theor. Biol.*, 203:229–248, 2000. 2, 4, 5, 12

[SMO+97]   E. Selkov, N. Maltsev, G.J. Olsen, R. Overbeek, and W.B. Whit-
            man. A reconstruction of the metabolism of *Methanococcus jan-
            naschii* from sequence data. *Gene*, 197:GC11–GC26, 1997. 106

[SP00]   C.H. Schilling and B.O. Palsson. Assessment of the metabolic capa-
            bilities of *Haemophilus influenzae* rd through a genome-scale path-
            way analysis. *J. theor. Biol.*, 203:249–283, 2000. 4

[SPA+02]   M. Steffen, A. Petti, J. Aach, P. D'haeseleer, and G. Church. Au-
            tomated modelling of signal transduction networks bmc. *Bioinfor-
            matics*, 3:34, 2002. 4, 10

[SPM+02]   S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar.
            Exploring the pathway structure of metabolism: decomposition into
            subnetworks and application to *Mycoplasma pneumoniae*. *Bioinfor-
            matics*, 18:351–361, 2002. 4, 30, 38, 64, 73

[Sta90]   P.H. Starke. *Analyse von Petri-Netz-Modellen*. B.G.Teubner,
            Stuttgart, 1990. 9, 54, 59, 61

[STKS00]   A.S. Sahota, J.A. Tischfield, N. Kamatani, and H.A. Sim-
            monds. Adenine phosphoribosyltransferase deficiency and 2,8-
            dihydroxyadenine lithiasis. In C.R. Scriver, W.S. Sly, B. Childs,
            A.L. Beaudet, D. Valle, K.W. Kinzler, and B. Vogelstein, editors,
            The Metabolic and Molecular Bases of Inherited Disease, volume 2,
            pages 2571–2584. McGraw-Hill, New York, 2000. 43

[Str95]   L. Stryer. *Biochemistry*. Freeman, New York, 1995. 4, 21, 36, 58,
            61, 66, 69, 70, 73, 104

[TB01]   J. Taipale and P. Beachy. The hedgedog and wnt signalling in can-
            cer. *Nature*, 411:349–354, 2001. 107

[Tom97]   I. Tomescu. *Data Structures*. Bucharest University Press,
            Bucharest, 1997. 79

[TTK95]   R. Thomas, D. Thieffry, and M. Kaufman. Dynamical behaviour
            of biological regulatory networks–i. biological role of feedback loops
            and practical use of the concept of the loop-characteristic state.
            Bull. Math. Biol., 57:247–76, 1995. 4, 10

[TWvDW98]   B. Teusink, M.C. Walsh, K. van Dam, and H.V. Westerhoff. The
            danger of metabolic pathways with turbo design. *Trends Biochem.
            Sci.*, 23:162–169, 1998. 1, 2, 10, 68, 104

[TYW96]   S. Tanimoto, M. Yamauchi, and T. Watanabe. Finding a minimal siphon containing specified places in a general petri net. In *IE-ICE Trans. on Fundamentals in Electronics, Communications and Computer Science*, volume E79-A, pages 1825–1828, 1996. 65

[UUY+00]   T. Ueki, T. Uyama, K. Yamamoto, K. Kanamori, and H. Michibata. Exclusive expression of transketolase in the vanadocytes of the vanadium-rich ascidian, *Ascidia sydneiensis samea. Biochim. Biophys. Acta.*, 1494:83–90, 2000. 17

[Var93]   K. Varpaaniemi. *Efficient Detection of Deadlock in Petri Nets.* Licentiate's thesis, Helsinki University of Technology, Department of Computer Science and Engineering, Digital Systems Laboratory, 1993. 66

[VdJ00]   G. Van denBerghe and J. Jaeken. Adenylosuccinate lyase deficiency. In C.R. Scriver, W.S. Sly, B. Childs, A.L. Beaudet, D. Valle, K.W. Kinzler, and B. Vogelstein, editors, *The Metabolic and Molecular Bases of Inherited Disease*, volume 2, pages 2653–2662. McGraw-Hill, New York, 2000. 45

[VDL02]   S.J. Van Dien and M.E. Lidstrom. Stoichiometric model for evaluating the metabolic capabilities of the facultative methylotroph *Methylobacterium extorquens* am1, with application to reconstruction of c(3) and c(4) metabolism. *Biotechnol. Bioeng.*, 78:296–312, 2002. 4, 5, 64, 104

[VHR+01]   N.M. Verhoeven, J.H.J. Huck, B. Roos, E.A. Struys, G.S. Salamons, A.C. Douwes, M.S. Van der Knaap, , and C. Jakobs. Transaldolase deficiency: liver cirrhosis associated with a new inborn error in pentose phosphate pathway. *Am. J. Hum. Genet.*, 68:1086–1092, 2001. 18

[WAL87]   J.F. Williams, K.K. Arora, and J.P. Longenecker. The pentose pathway: a random harvest. *Int. J. Biochem.*, 19:749–817, 1987. 4, 21, 23

[WNT+74]   Y. Wada, Y. Nishimura, M. Tanabu, Y. Yoshimura, K. Iinuma, T. Yoshida, and T. Arakawa. Hypouricemic, mentally retarded infant with a defect of 5-phosphoribosyl-1-pyrophosphate synthetase of erythrocytes. *Tohoku J. Exp. Med.*, 113(2):149–157, 1974. 49

[WPS+00]   J. Wolf, J. Passarge, O.J.G. Somsen, R. Snoep, J.L.and Heinrich, and H.V. Westerhoff. Transduction of intracellular and intercellular dynamics in yeast glycolitic oscilations. *Biophys. J.*, 78:1145–1153, 2000. 1, 2

[WSL03]   H.S. Wiley, S.Y. Shvartsman, and D.A. Lauffenburger. Computa-
           tional modeling of the egf-receptor system: a paradigm for systems
           biology. *Trends Cell Biol.*, 13(2):43–50, 2003. 4, 10

[YG84]   B. Younkin and B. Gudzinowicz. The viral mechanism of reye's
           syndrome. *Med. Hypotheses*, 14(2):161–80, 1984. 50

[YO80]   M. Yudkin and R. Offord. *A Guidebook to Biochemistry*. University
           Press, Cambridge, 1980. 21

[YTW96]   M. Yamauchi, S. Tanimoto, and T. Watanabe. Finding a minimal
           siphon containing specified places in a general petri net. In *IE-
           ICE Trans. on Fundamentals in Electronics, Communications and
           Computer Science*, volume E79-A, pages 1825–1828, 1996. 65

[YW99]   M. Yamauchi and T. Watanabe. Time complexity analysis of the
           minimal siphon extraction problem of petri nets. In *IEICE Trans.
           on Fundamentals in Electronics, Communications and Computer
           Science*, volume E82-A, pages 2558–2565, 1999. 65

[ZHM+02]   L. Zalah, M. Huleihel, E. Manor, A. Konson, H. Ford, V.E. Mar-
           quez, D.G. Johns, and R. Agbaria. Metabolic pathways of n-
           methanocarbathymine, a novel antiviral agent, in native and herpes
           simplex virus type 1 invefted vero cell. *Antiviral Research*, 55:63–75,
           2002. 51

Acknowledgements

Lebenslauf

Nachname: Oancea (geb. Zevedei) *Vorname*: Ionela

Adresse: Groscurthstr. 28, 13125, Berlin, Deutschland

Bd. Iuliu Maniu nr. 74-76, Bl. 5, Sc. 4, Ap. 141, Sector 6, Bukarest, Rumänien.

Geburtsdatum und -ort: 22. Februar 1977, Bukarest, Rumänien

Ausbildung

Abitur:

Datum: Juni 1995

Ort: Lyzeum G. Lazar, Bukarest (entspricht der gymnasialen Oberstufe)

Bemerkung: Abschlussnote 9,87

(im rumänischen System laufen die Noten von 10[sehr gut] bis 1)

Diplom:

Datum: Juni 1999

Ort: Fakultät für Mathematik, Universität zu Bukarest

Fachgebiet: Informatik

Betreuer: Dr. L Popescu

Diplomarbeitsthema: Elementar-Codes

Bemerkung: Abschlussnote 9,50

Master in Informatik:

Datum: Februar 2001

Ort: Fakultät für Mathematik, Universität zu Bukarest

Fachgebiet: Informatik

Spezialisierung: Betriebssysteme und parallele Verfahren

Betreuer: Prof. Dr. I. Vaduva

Thema der Masterarbeit: Programmierung einer graphischen Schnittstelle für ein datenbankgestütztes Durchführbarkeits-Management System (Programmierung eines Grafikinterfaces zu einer Datenbank für Machbarkeitsstudien)

Bemerkung: Abschlussnote 9,90

Promotionsstudium / Doktorandenstipendium:

Datum: Beginn Juni 2000

Ort: Graduiertenkolleg „Dynamik und Evolution zellulärer und makromolekularer Prozesse", Institut für Biologie, Humboldt-Universität zu Berlin

Fach: Biophysis

Betreuer: PD. Dr. S. Schuster

Veröffentlichungsliste

Schuster S., Zevedei-Oancea I. (2002) Treatment of multifunctional enzymes in metabolic pathway analysis. *Biophys. Chem.* **99**:63-75.

Zevedei-Oancea I., Schuster S. (2003) Topological analysis of metabolic networks based on Petri net theory. *In Silico Biology* **3**, 0029 (on-line, Papierversion im Druck).

Zevedei-Oancea I., Schuster S., SigNetRouter - Framework for detecting signal transfer routes in signalling networks eingereicht bei Computers & Chemical Engineering).

Bolohan Orest, Oancea Cristian, Zevedei Ionela, Santa Ionel (2002) SQL- Data manipulation language in "Procedural and non-procedural query solving in Oracle 8", Popescu Ileana, *University Press*, Bucharest, 33-48.

Erklärung

Ich versichere hiermit, die vorliegende Arbeit selbständig und ausschließlich unter Verwendung der angegeben Mittel und ohne unerlaubte Hilfen angefertigt zu haben.

Ionela Oancea   Berlin, den 15. Juli 2003