VRIJE UNIVERSITEIT

Of Molecules and Cells: emergent mechanisms

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. T. Sminia,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Aard- en Levenswetenschappen
op vrijdag 1 juli 2005 om 10.45 uur
in de aula van de Universiteit,
De Boelelaan 1105

door

Frans Johannes Bruggeman

geboren te Dirksland

Printed by Febo druk BV, Enschede

1. From Molecular Biology to Systems Biology

The 20th century witnessed the growth of biology into a major interdisciplinary field. Especially, the elucidation of the molecular organization of living cells serves as an example: *e.g.* the discovery of enzymes (cf. Chapter 7 in Bechtel & Richardson (23))[1], of metabolic pathways (the elucidation of glycolysis was finished by the end of 1930's (23)), of compartmentation (organelles), of allosteric regulation (290, 389), of the structure of DNA (403), of the proton motive force (264), of the central dogma of molecular biology (cf. 402), and the determination of the DNA sequences of many organisms, including man (396). These are all examples of major biological and scientific achievements that have had, and are anticipated to continue to have, large consequences for technology, society, and for our understanding of nature. At present, the accumulated knowledge of the properties of the macromolecules that make up living cells is daunting in magnitude and continues to grow at an enormous pace.

Nowadays, it has become relatively straightforward to obtain organism-wide data sets (X-omics data sets) of the levels of mRNAs (transcriptomics), proteins (proteomics), metabolites (metabolomics), and fluxes (fluxomics) for populations of cells growing at defined and controlled conditions (*e.g.* (13, 30, 54, 74, 86, 99, 115, 126)). With such methods a large proportion of the average state of a population of cells can be assessed. The simultaneous determination of those data sets should enable the determination of the cell's state almost in its entirety. The integration of X-omics data sets with computational and theoretical methods is anticipated to become important (153, 158, 211, 415). Technology that will facilitate the analysis of the spatial organization of cells by analysis of the (spatiotemporal) states of single cells is being developed actively (60, 169, 170, 220, 252). All of this illustrates that the developments in contemporary biology have become highly dependent on technological innovations.

The wealth of biological data that is now being generated is being stored in enormous databases. They contain, for example: the DNA sequence of entire organisms, the functional annotation of their genes, the three-dimensional structure of many of their proteins, and the structure of their biochemical reaction networks (metabolic pathways, signaling networks, and genetic networks). These databases continue to grow in number and in sophistication (122). Nowadays we can compare entire organisms regarding the organization of their genome on the basis of genome-sequence homology, *i.e.* which proteins are likely to be present or not.

[1] In the first decade of 1900, most biochemical research was done in yeast – 'enzyme' means 'in yeast'.

This facilitates the analysis of the evolutionary history of and relationships among species. In addition, it allows for the formulation of a hypothesis on the stoichiometric structure of the biochemical reaction network of sequenced organisms, of which only a limited amount of biochemical data, or none at all, is available, on basis of the comparison of its unannotated genome with the annotated genomes of related organisms. Bioinformatics is the branch of biology that maintains and analyses these databases and has become an extremely data-rich and promising discipline.

Despite the enormous achievements in molecular biology, genetics biochemistry, and cellular biology, precise manipulation of living cells for medical, technological, or scientific purposes remains problematic. These three disciplines have in common that their focus is intralevel, *i.e.* on either the behavior of the components of cells or the behavior of cells as a whole. At the present time, many scientists argue that the integration of these disciplines – to form an *interlevel* systems biology – should enable biology to take the next step forward (see the special issue on systems biology in *Science* (1th of March issue of Science, 295, 2002) and in *Nature* (14th of November issue of Nature, 420, 2002) and (13, 153, 158, 211, 415)).

The need for a systems biology approach becomes evident from an example taken from oncology. Cancerous cells display common characteristics mainly at the *systemic* cellular level, *i.e.* dysregulated growth of cells. These systemic aberrations are due to the malfunctioning of a number of cellular *components*, (proteins), which ultimately results from a number of mutations in the genome of this particular cell induced by, for instance, UV-radiation or smoking. To be able to restore 'healthy' cellular functioning, ideally a number of specific processes have to be targeted by the administration of drugs (inhibitors or activators of cellular processes). In most practical cases, however, it is not known to what extent particular processes contribute to systemic functioning in cancerous cells and therefore it is not clear which processes have to be targeted. (Accordingly, present day cancer therapy focuses for a large part on the elimination of the cancerous cells.) Metabolic control analysis (MCA) has illustrated that the control of particular processes on systemic properties is likely to be distributed over many different processes and that the control distribution changes with the state of the cell (134, 179); for an illustration of the application of MCA to eukaryotic signal transduction see Hornberg *et al.* (154, 155) and Krauss *et al.* (226, 227). MCA indicates that qualitative data on the importance of a particular processes for systemic functioning rarely suffices to identify drug targets (34, 259). Cell biology appears to be in need of a method that links the behavior of macromolecules to the behavior of cells in a quantitative manner. A critical component in this method seems to be the development and

analysis of detailed kinetic models of signaling networks.

The disciplines that analyse macromolecular and cellular behavior have classically been biochemistry and physiology, respectively. These disciplines were once very large, but they have lost terrain after the molecular-genetic revolution that eventually led to the rise of the molecular biology and, more recently, of X-omics (*i.e.* metabolomics, transcriptomics, proteomics, fluxomics). At the same time, advances in biophysics, immunology, and microscopy have pushed cell physiology more towards molecular-cell biology; where the relation between the cell and its constituent molecules is the focus. At present, the research into networks of interacting macromolecules as a bridge between macromolecular and cellular is becoming a promising field under the name of 'Systems Biology' (158, 211, 409). Systems biology investigates how the behavior of cells is brought about by the interactions and spatiotemporal organization of their molecular constituents and, therefore, its emphasis differs from that of biochemistry (which sought function in individual macromolecules) and physiology (which describes function without referring to the molecules basis). Thus systems biology aims at becoming an *interlevel* science that connects two levels of cell-biological organization, *i.e.* the cell and its constituents. In contrast, biochemistry and physiology appear more *intralevel* in their focus.

The nonlinearity of the cell, considered as a system of organized and interacting (macro)molecules, makes the use of mathematical models of biochemical networks practically a necessity (see Section 2.8 for an example). Mathematical models constructed on the basis of the accurately measured kinetic and physicochemical properties of the macromolecules and the structure of the network, *i.e.* so-called silicon cells (409), aim at being as precise replicas of cellular networks as possible. They 'operationalize' molecular-biological and biochemical knowledge conform the laws of physics and allow for the calculation of the systemic consequences of particular macromolecular properties in the context of the networking with other macromolecules; they attempt to bridge the gap between molecules and cells. Subsequently, the model predictions can be tested in the laboratory; lack of correspondence then points at incomplete molecular or organizational understanding and need for model improvement. The iteration between mathematical-model predictions, subsequent testing in the laboratory, and model refinement is a promising paradigm that has proven to be successful in physics and chemistry and has now become within reach of cell biology.

2. The description of cellular networks *in silico*

2.1 Intracellular organization

Many mathematical models of cells attempt to reconstruct the behavior of macromolecules *in vivo*. Hereby they assist in describing and even analyzing the cellular mechanisms at work (9, 137, 387, 388, 407). However, the construction of quantitative mathematical models of cellular phenomena requires a critical evaluation of the organization of the interior of cells. The cellular milieu in which the interactions among the molecules take place is likely to be far from the conditions generally applied *in vitro* (93) (cf. Figure 1; Table 1). One of the more obvious determinants of the complicated interior of an eukaryotic cell is its spatial organization, *e.g.* arising from compartmentalization in organelles or from diffusion limitation. A dynamic spatial organization (*e.g.* waves) may result from diffusion gradients and nonlinear processes (186). Macromolecular crowding may lead to decreased diffusion coefficients of macromolecules as compared to their values in more diluted environment as frequently used *in vitro* (94). Macromolecular crowding can also influence the values of rate constants and, concomitantly, of dissociation and association constants (92, 130, 263, 324, 341, 385, 433). Concentration gradients caused by low diffusion rates could lead to spatiotemporal structure formation including cytosolic waves of intermediates in signaling networks (42, 186). Confinement of processes to areas adjacent to membranes may enhance association equilibria (193). Macromolecular crowding may also enhance the occurrence of channelling: the direct transfer of catalytic products by donor enzymes to acceptor enzymes as substrates – in many cases these enzymes are part of large enzyme complexes (286, 324).

Another source of complexity in the organization of cells is the stochastic nature of many intracellular processes. The occurrence of low molecule numbers or severe nonlinearity may increase the magnitude of fluctuations in particle numbers and increase the significance of stochastic behavior of processes (8, 88-90, 251). Recently, it was shown that biochemical networks that operate in regimes with zero-order kinetics may display large fluctuations even if the copy numbers of the individual molecular species are on the order of thousands (88, 303). This illustrates that the generally applied rule of thumb that for processes removed from thermodynamic equilibrium fluctuations in the concentration of an intermediate x are approximately of the order of the inverse of the square root of its average (macroscopic) copy number $\langle x \rangle$, *i.e.* $1/\sqrt{\langle x \rangle}$, has only limited validity. Stochastic fluctuations may be particularly important in systems that are close to criticality (a saddle-node bifurcation) or in signaling networks that contain zero-order ultrasensitive signaling cycles (*e.g.* 90, 301).

The aforementioned intracellular phenomena may influence each other *in vivo* in non-intuitive ways. This may result in differences between the dynamics of cellular networks *in vivo* and the networks reconstructed *in silico* on the bases of *in vitro* kinetic data. On the other hand, there are many examples where *in vitro* kinetic data proved to suffice in models that did not include diffusion, crowding, and stochastic descriptions (*e.g.* 17, 146, 378). Only in some cases, discrepancies between *in vivo* and *in silico* behavior necessitate the use of more complicated modeling approaches. For instance, low diffusion coefficients may necessitate the use of partial differential equations (reaction-diffusion equations (*e.g.* 111)) whereas low particle numbers or the possibility of critical behavior may necessitate usage of the master equation for the description of stochastic phenomena (*e.g.* 8, 89, 90). Other approaches include particle-based modeling and hybrid modeling (combining ordinary, partial, and stochastic differential equations). Another possibility for discrepancies between *in vivo* and *in silico* behavior is that the kinetic data have to be determined *in vivo* rather than *in vitro*. In Figure 2 the differences of various modeling descriptions on the dynamics of metabolic systems is illustrated with an example.

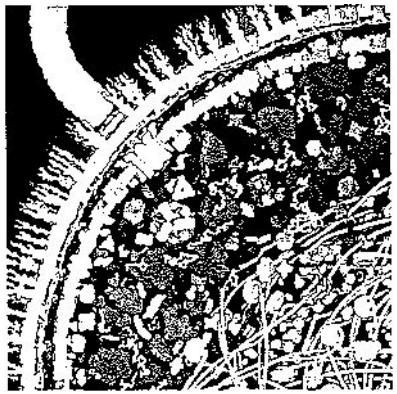

Figure 1.

An artist impression of E. coli based on the relative dimensions of its macromolecular constituents (123). The cell wall is displayed (harboring a protruding flagellum), which surrounds the cytosol. The cytosol contains proteins, ribosomes, mRNAs, DNA (coiled around bacterial nucleosomes), and DNA polymerases. Used with permission from David Goodsell (http://www.scripps.edu/pub/goodsell/).

2.2 Intracellular processes

Not only the type of modeling description, *e.g.* with/without diffusion or with/without stochasticity, is to be decided upon, but also the level of biochemical detail has to be chosen (Figure 2 & 3). The latter aspect of network model construction is not entirely independent of what was discussed in the previous section, but for simplicity we will discuss this separately. The level of biochemical detail depends to a large extent on the time scale of interest, the quantitative nature of the question, and the quality and the amount of the experimental data available. For instance, if the dynamics of metabolites upon a perturbation of metabolism is of interest, enzymes can be generally approximated with quasi-steady state kinetics, *e.g.* by Michaelis-Menten type rate equations, provided certain dynamic criteria are met (*cf.* (216, 321, 353)). On the other hand, if shorter time scales are of interest, the intra-enzymatic conversions should be considered; that is, the dynamics of enzyme-substrate complexes themselves (353).

Table 1. *Characteristics of* Escherichia coli.

Characteristic	*Quantity*
Shape	Cylinder with spherical poles
Diameter	0.5-1.5 μm (1)
Length	2-4 μm (1)
Volume/Area	2.9 μm^3/12.6 μm^2
Mass of one cell	1-2 pg (70% water and 30% biomass) (2)
Diffusion constant	Protein: 3.6-7.7 $\mu m^2 s^{-1}$ / Glucose: 670 $\mu m^2 s^{-1}$ (1)
Genome size	$4.6*10^6$ bp (4377 genes: 4290 proteins rest RNAs) (3)
Kinds of molecules	1500 (2)
Macromolecules (% of total drw)	96.1 (2) (4)
Soluble pool (% drw)	2.9 (2)
Inorganic ions (% of total drw)	1 (2)
Doubling time	20-60 minutes

1. (111)
2. (162) (growth conditions: mineral medium, glucose, doubling time 40 minutes).
3. http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/G/GenomeSizes.html
4. protein/RNA/DNA/Lipid/Lipid/polysaccharide/Peptidoglycan/Glycogen=55/20.5/3.1/9.1/3.4/2.5/2.5 (162)

There exist also models that do not contain any kinetic information on the processes but solely contain the stoichiometry of the reactions that make up the network. (For instance, for the reaction $2A \leftrightarrow A_2$ the stoichiometric coefficient of A and A_2 in the direction of synthesis of A_2 is -2 and 1, respectively.) Such models will be referred to as stoichiometric

models (Figure 3). Stoichiometric models are mostly constructed for analysis of the steady-state flux properties of metabolic networks to address problems related to, for instance; the flux distribution (374), the flexibility and robustness of metabolism (368), the flux distribution with optimal yield of some desired product (344), the identification of minimal cut sets (214) and enzyme subsets (294), or the viability of mutarts (345). (For a recent application to signaling networks see (100).)

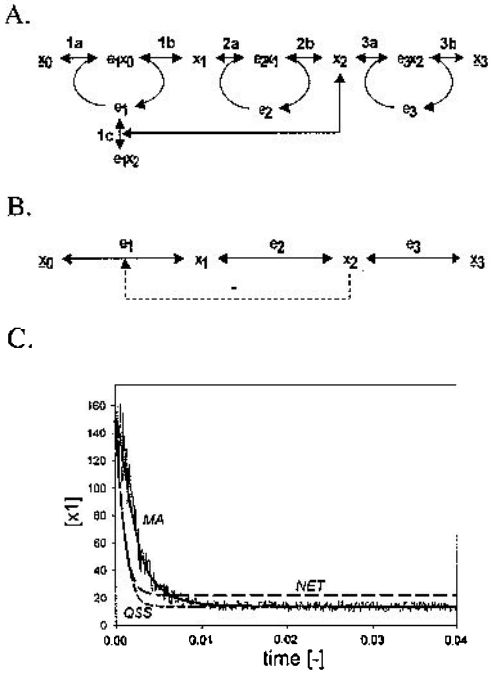

Figure 2.

Different modeling methods and descriptions illustrated for one metabolic pathway. Figure A and B depict the same pathway but with different levels of coarse graining. In A, the individual binding events are considered and described by mass action kinetics (MA). In B, the enzymes are treated as wholes described by quasi-steady state enzyme kinetics (QSS). In C, the differences in the dynamics of x_1 is shown between: (i) the mesoscopic description in terms of the master equation (116) and the macroscopic description in terms of ode's using mass action kinetics (e.g. 137), and (ii) the macroscopic descriptions in terms of quasi-steady state (qss) enzyme kinetics (353) and a lin-log approximation (NET) of the enzyme rate equations. The lin-log approximation is based on description of rates of enzymes by irreversible thermodynamics (397, 407). The mesoscopic description is the fluctuating line around the solution obtained with mass-action kinetics. Note that MA and QSS obtain the same steady state for x_1 but that NET describes a different steady state value.

Roughly, two types of kinetic models can be discerned (Figure 3). The first type is the core model. Core models have both simplified kinetics and simplified stoichiometric structures, *e.g.* lumped enzyme reactions. Examples of simplified kinetics are simplified rate equations (*e.g.* irreversible, product independent, simplified mechanism, etc.) and educated guesses for kinetic constants. This means that these models are less detailed and that they mimic *in vivo* behavior qualitatively at best. Their advantage is that they are easy to analyze and that they are small. The second type of kinetic model is the detailed model with both exactly determined stoichiometric structure and kinetic properties, *i.e.* a silicon cell (www.jjj.bio.vu.nl). Glycolytic oscillations provide an interesting example of a biological phenomenon where both core models (121, 354) and detailed models (157, 310) have been used. Table 2 compares the different models of cellular networks according to five criteria. Clearly, the detailed model is the most complete description but, therefore, also hard to analyze to obtain specific explanations for behavior.

Stoichiometric and kinetic analysis of cellular networks is combined in the analysis of the dynamics and the control of cellular networks as will be shown in some detail in the next sections.

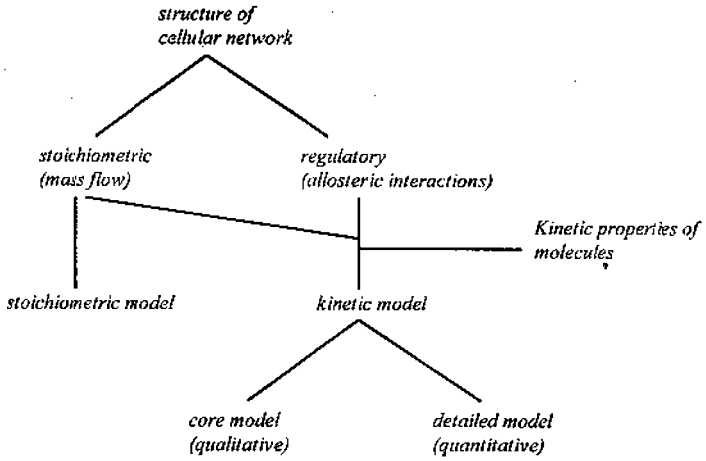

Figure 3.

Different kinds of models can be constructed for cellular networks. The model descriptions differ with respect the type of information on the cellular network and the level of detail that is used. This scheme could be expanded by including models take into account additionally, the spatial organization of the cellular network macromolecular crowding and/or the stochastic nature of processes, but, for reasons of simplicity, this is not shown.

Table 2. *Comparison of three different types of models of cellular networks.*

Model type	Quantitative	Dynamics	Regulation	Kinetics	Analyzability
Stoichiometric	+	-	-*	-	+
Core	-	+	+	+/-	+
Detailed	+	+	+	+	

* for an exception see (67) & (241).

2.3. Dynamics of biochemical reaction networks

The networks considered here will be described in terms of autonomous ordinary differential equations (ode's); that is, it will be assumed that mass transport by diffusion is sufficiently fast to prevent the development of concentration gradients and that fluctuations in concentrations are too small to determine systemic behavior. In addition, we will assume that the volume of the cellular system remains constant over time (for a relaxation of this assumption see (321)). We will use the following notation: scalars, vectors and matrices will be denoted by normal face, bold normal face and bold capital letters, respectively – with the flux vector **J** as an exception.

The dynamics of a cellular network composed of r reactions (*e.g.* membrane transport, dissociation or association events, catalytic conversions, covalent modifications) and m variable intermediates (*e.g.* metabolites, enzymes, transcription factors, mRNA's) can be captured in terms of m ode's that describe the rates of change of the concentrations of the m variable species at time t and constitute the mass balances (137):

$$\frac{d}{dt}\mathbf{s}(t,\mathbf{p}) = \mathbf{F}(\mathbf{s}(t,\mathbf{p}),\mathbf{p}) = \mathbf{N}\cdot\mathbf{v}(\mathbf{s}(t,\mathbf{p}),\mathbf{p}) \qquad (1)$$

The $mx1$ vector $\mathbf{s}(t,\mathbf{p})$ denotes the concentrations of all the molecules with variable concentrations at time t: it represents the state of the cellular network at time t. It additionally depends on parameters contained in the vector $\mathbf{p}$ that characterize: (i) the kinetic and thermodynamic properties of the molecular constituents (*e.g.* half-saturation constants (K_M), maximal rates (V_{MAX}), equilibrium constants (K_{EQ}), dissociation constants (K_D)), (ii) the environment (constant (intra- or extracellular) concentrations, temperature), and (iii) the concentrations of the variable intermediates at time zero ($\mathbf{s}_0=\mathbf{s}(0,\mathbf{p})$). The mxr stoichiometric matrix $\mathbf{N}$ contains the stoichiometry of the cellular network, *e.g.* a positive (or negative) (i,j)-th entry of $\mathbf{N}$ gives the number of molecules s_i produced (or consumed) by the j-th reaction v_j of the $rx1$ rate vector v. The stoichiometric matrix is identical to the matrix $\partial\dot{\mathbf{s}}/\partial\mathbf{v}$ (with $\dot{\mathbf{s}} = d\mathbf{s}/dt$). The changes in the rates at time t can be obtained from, $\dot{\mathbf{v}} = \partial\mathbf{v}/\partial\mathbf{s}\,\dot{\mathbf{s}}$.

Many metabolites take part in moiety-conservation relationships. These relationships are linear combinations of concentrations of metabolites that remain constant within a particular time window and generally involve metabolites that have no net transport in or out of the cell, *e.g.* the total phosphate concentration P_T on a particular time scale can, for instance, obey $P_T=Pi+AMP+2ADP+3ATP$ (assuming no import and export of Pi). In some cases, moiety-conservation relationships between metabolites are readily recognized and may be taken into account by resolving to descriptions in terms of attractive independent variables such as the ATP/ADP or NADH/NAD ratios (416). In complicated networks it is often hard to recognize the moiety-conservation relationships and a more systematic approach is necessary. These relationships are reflected by linear dependencies between the rows of the stoichiometric matrix **N** (309). The number of independent metabolites equals the rank of the stoichiometric **N** and is denoted by m_0. Linearly-dependent rows can be identified by row reduction into echelon form (*cf.* (63)), where software packages such as *Mathematica*, *Maple*, *Gepasi* (256, 257), *Jdesigner* or *Jarnac* can be of assistance. Each moiety-conservation relationship makes the concentration of one metabolite dependent on the concentrations of other metabolites. Thus the number of dependent metabolite concentrations equals the number of moiety-conservation relationships, *i.e.* m-m_0. The metabolite vector can be decomposed into two subvectors: a subvector that contains independent variable concentrations and a subvector that contains dependent variables, *i.e.* $\mathbf{x}$ and $\mathbf{x}^D$, respectively, in other words, $\mathbf{s}=(\mathbf{x},\mathbf{x}^D)^T$ (with superscript "T" denoting the transpose of a matrix or a vector). The decomposition of $\mathbf{s}$ can be paralleled by a decomposition of $\mathbf{N}$ into $[\mathbf{N}^0;\mathbf{N}']^T$, which has the following consequences for the description of the dynamics of the network (309):

$$\frac{d}{dt}\mathbf{x}(t,\mathbf{p}) = \mathbf{N}^0\mathbf{v}(\mathbf{x}(t,\mathbf{p}),\mathbf{x}^D(\mathbf{x}(t,\mathbf{p})),\mathbf{p}) \qquad (2a)$$

$$\frac{d}{dt}\mathbf{s} = \frac{d}{dt}\begin{bmatrix}\mathbf{x}\\ \mathbf{x}^D\end{bmatrix} = \begin{bmatrix}\mathbf{I}\\ \frac{\partial\mathbf{x}^D}{\partial\mathbf{x}}\end{bmatrix}\frac{d}{dt}\mathbf{x} \equiv \begin{bmatrix}\mathbf{I}\\ \mathbf{L}'\end{bmatrix}\frac{d}{dt}\mathbf{x} \equiv \mathbf{L}\frac{d}{dt}\mathbf{x} = \mathbf{L}\mathbf{N}^0\mathbf{v} = \mathbf{N}\mathbf{v} \qquad (2b)$$

These equations indicate that the dynamics of the dependent intermediates can be obtained from the dynamics of independent intermediates. A simple method to compute the L matrix is from (137, 334):

$$\mathbf{L}'\cdot\mathbf{N}^0 = \mathbf{N}' \quad \Rightarrow \quad \begin{bmatrix}-\mathbf{L}' & \mathbf{I}\end{bmatrix}\cdot\begin{bmatrix}\mathbf{N}^0\\ \mathbf{N}'\end{bmatrix} = \mathbf{0}$$

The matrix $[-\mathbf{L}'\ \ \mathbf{I}]$ is defined as the conservation matrix G and so we obtain:

$$\mathbf{GN} = \mathbf{0} \ \Rightarrow\ \mathbf{N}^T\mathbf{G}^T = \mathbf{0}$$

Which means that L can be obtained from the right nullspace of $\mathbf{N}^T$ or from the left nullspace of N. Alternative methods for determining the L matrix can be found in Sauro & Ingalls (334).

2.4. Description of rates of processes

Entries of the rate vector v are equations that characterize the dependency of a rate of a particular process on the concentration of the enzyme, on the concentrations of the substrates, the products, and the effector molecules, and on the kinetic parameters of the process. For instance, the rate of enzyme j that catalyzes a single-substrate (concentration: S) single-product (concentration: P) reaction v_j and which is inhibited by a noncompetitive effector (concentration: I) can be described by a reversible product-dependent rate equation such as (*cf.* (353)):

$$v_j = \frac{V_j^+ \frac{S}{K_{M,S,j}} - V_j^- \frac{P}{K_{M,P,j}}}{\left(1+\frac{I}{K_{M,I,j}}\right)\left(1+\frac{S}{K_{M,S,j}}+\frac{P}{K_{M,P,j}}\right)} = \frac{V_j^+ \frac{S}{K_{M,S,j}}\left(1-\frac{P/S}{K_{EQ,j}}\right)}{\left(1+\frac{I}{K_{M,I,j}}\right)\left(1+\frac{S}{K_{M,S,j}}+\frac{P}{K_{M,P,j}}\right)} \qquad (3)$$

The kinetic parameters of this reaction are the half-saturation constants $K_{M,S,j}$, $K_{M,P,j}$, and $K_{M,I,j}$ of the enzyme for S, P, and I, respectively, the maximal rate in the forward and backward direction, respectively, *i.e.* V_j^+ (the product of a catalytic rate constant and the concentration of the enzyme: $k_{cat,j}^+ \cdot e_j$) and V_j^- ($k_{cat,j}^- \cdot e_j$), and the equilibrium constant of the reaction $K_{EQ,j}$.

The Haldane relationship relates the kinetic parameters of the enzyme to the equilibrium constant, which is determined by the thermodynamic properties of the substrate and product molecules (cf. 56, 353):

$$K_{EQ,j} = \frac{V_j^+ K_{M,P,j}}{V_j^- K_{M,S,j}} \qquad (4)$$

The rate of an enzymatic reaction is connected to the thermodynamic properties of reactants through the mass-action ratio ($\Gamma_j = P/S$) and the chemical potential of the reaction ($\Delta\mu_j$).

For a one substrate and one product reaction we obtain:

$$\Delta\mu_j = RT\ln\frac{P}{S\cdot K_{EQ,j}} = RT\ln\frac{\Gamma_j}{K_{EQ,j}}$$

This equation indicates that if the reaction is in thermodynamic equilibrium the chemical potential is zero and that its rate will also equal zero (see Eq 3) (407). The displacement from thermodynamic equilibrium ($1-\Gamma_j/K_{EQ,j}$) appears in the equations for the (scaled) elasticities of enzymes towards their substrates and products (102, 407):

$$\varepsilon_S^v \equiv \left(\frac{\partial \ln v}{\partial \ln[S]}\right)_{[P]} = \frac{1}{1-\frac{\Gamma}{K_{EQ}}} - \frac{\frac{S}{K_S}}{1+\frac{S}{K_S}+\frac{P}{K_P}}$$

$$\varepsilon_P^v \equiv \left(\frac{\partial \ln v}{\partial \ln[P]}\right)_{[S]} = \frac{-\frac{\Gamma}{K_{EQ}}}{1-\frac{\Gamma}{K_{EQ}}} - \frac{\frac{P}{K_P}}{1+\frac{S}{K_S}+\frac{P}{K_P}}$$

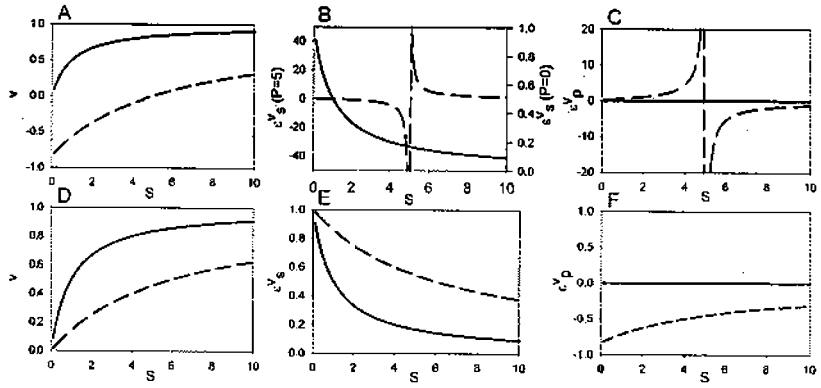

Figure 4.

The upper and lower figures respectively depict the rates (A, D) and elasticities to substrate (B, E) and product (C, F) for a reversible and an irreversible enzyme following an uni-uni Michaelis-Menten kinetic mechanism. Parameters: Ks=1, Kp=1, V=1, and Keq=10 (reversible case). Dashed lines correspond to a value for P of 5 and full lines correspond to a value for P of 0.

The total elasticity of the rate of the enzyme towards the metabolite environment decreases with the total saturation of the enzyme:

$$\varepsilon_S^v + \varepsilon_P^v = 1 - \frac{\frac{S}{K_S} + \frac{P}{K_P}}{1 + \frac{S}{K_S} + \frac{P}{K_P}}$$

If the saturation term is negligible (if $S \ll K_S$ and $P \ll K_P$ or when the saturation term is entirely absent in case of mass-action kinetics) the sum of the elasticities is 1, which indicates that the rate increases proportionally with a simultaneous equal fractional change in the concentration of S and P. If the enzyme is saturated ($S \gg K_S$ and $P \gg K_P$) the total elasticity converges to zero although the two individual elasticities do not. The rates and the elasticities of an enzyme modelled as Eq. 3 without the inhibition term is depicted in Figure 4.

In the sixties and seventies of the previous century, the determination of the mechanisms by which enzymes convert their substrates into products and how the catalytic rate depends on the concentration of effectors and physicochemical conditions was a large field of research within enzymology (cf. 57, 58, 59, 223, 266, 353). Unfortunately, considering the present need for kinetic models, the determination of kinetic properties of enzymes *per se* was not considered important in those days. Enzymologists determined just as many kinetic properties as necessary to elucidate the catalytic mechanism of an enzyme. Also, the conditions under which the mechanism was determined, using purified enzymes or enzymes in diluted cell free extract, were mostly far from the conditions in the cytoplasm of cells, *e.g.* different pH, ionic strength, and extent of macromolecular crowding were used. In many cases, the mechanisms were found to be random – the substrates and products did not bind in a particular sequence to the enzyme – and frequently those mechanisms were approximated by more simple mechanisms that are 'preferred' by the enzyme. Random mechanisms are notorious for featuring large numbers of kinetic parameters (353). The kinetics of enzymes with complicated mechanisms are frequently simplified in computer models to ordered quasi-steady state (207) or equilibrium-binding mechanisms (353). The kinetics of *multimeric* enzymes is further complicated by the possibility of cooperative behavior. Cooperative behavior stands for the phenomenon that the kinetic properties of a particular subunit of a multimeric enzyme may depend on the binding state of one or more other subunits of the same enzyme. Cooperative behavior may lead to complicated dependencies of the rates of multimeric enzymes on the concentrations of their substrates, products, and effectors. Whereas these dependencies are mostly hyperbolic for monomeric

enzymes, they may show maxima and minima for multimeric enzymes. The rate equations of multimeric enzymes are also very complicated and incorporate many rate constants (223, 266, 318, 353). So far, the kinetics of multimeric enzymes have been mostly investigated theoretically for multimeric enzymes catalyzing irreversible single substrate and single product reactions. Reversible reactions have been considered by Hofmeyr (149) and Popova (297). Popova (297) considered in addition descriptions of cooperative enzymes catalyzing multi-substrate and multi-product reactions.

2.5. Steady states of biochemical reaction networks

Cellular networks settle either to a point, a periodic, or a strange (chaotic) attractor (*cf.* (128)). Here we will focus on point attractors. A point attractor is a steady state, in which $\mathbf{s} = \bar{\mathbf{s}}$, if, after some time of relaxation, the following holds (where "≡" denotes definition):

$$\mathbf{N}^0\mathbf{v}(\bar{\mathbf{s}}(t,\mathbf{p}),\mathbf{p}) = \mathbf{0} \qquad (5a)$$

$$\mathbf{v}(\bar{\mathbf{s}}(t,\mathbf{p}),\mathbf{p}) \equiv \mathbf{J}(\bar{\mathbf{s}}(t,\mathbf{p}),\mathbf{p}) \text{ with at least one } J_i \neq 0 \qquad (5b)$$

The system is in thermodynamic equilibrium with $s=s_{EQ}$ if $\mathbf{v}(\mathbf{s}_{EQ}(t,\mathbf{p}),\mathbf{p}) = \mathbf{0}$ for all $v_i \in \mathbf{v}$. (Note that in mathematics no distinction is made between a steady state and an equilibrium state.) The steady-state rate vector, $\mathbf{v}(\bar{\mathbf{s}})$, will be denoted by the flux vector J.

The flux vector is a linear combination of the columns of the nullspace or kernel of N denoted by K: the columns of K span the steady-state flux space. The number of columns of K equals the number of $r\text{-}m_0$ independent fluxes of the network (each equation in Eq. 5 makes one flux dependent on other fluxes within the same equation). The following relationships hold:

$$\mathbf{NK} = \mathbf{0} \qquad (6a)$$

$$\mathbf{N}^0\mathbf{K} = \mathbf{0} \qquad (6b)$$

Reordering the rows of the flux vector with the independent fluxes $\mathbf{J}^I$ on top yields the following equation (with $\mathbf{J}^D$ as the dependent fluxes) (309):

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}^I \\ \mathbf{J}^D \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{K}' \end{bmatrix} \mathbf{J}^I = \mathbf{K}\mathbf{J}^I \qquad (7)$$

This equation indicates that all fluxes can be written as a linear combination of the independent fluxes. The columns of the kernel matrix are a non-unique set of independent pathways, so-called flux modes, that are each in steady state (or thermodynamic equilibrium) if the entire system is in steady state – not taking into account any kinetic information of the

processes involved. This only means that if all but one of the pathways corresponding to columns of the kernel would be removed from the network that this one pathway could attain a steady state on its own but it will have different values for the fluxes and the concentrations of its variable intermediates than in the original system. A zero entry in the kernel matrix refers to a reaction that is in equilibrium while the entire system is in a steady state (137). Unique sets of independent steady-state flux modes can be obtained by the determination of extreme pathways (55, 288, 339, 340) or elementary modes (294, 344-346, 400). Both methods take, besides the kernel matrix, the thermodynamically preferred direction of the reactions into account (*e.g.* on basis of values for equilibrium constants). Elementary flux modes are minimal sets of enzymes that can each generate valid steady states with all irreversible reactions proceeding in the direction prescribed thermodynamically, *i.e.* on basis of equilibrium constants. An elementary flux mode is elementary if it is non-decomposable. Any steady-state flux pattern can be expressed as a non-negative combination of the elementary flux modes (345). The number of extreme pathways is the minimal number of independent steady-state pathways, which means that, in some cases, there are more elementary modes than extreme pathways (288). For an in depth comparison of these two methods the reader is referred to Klammt *et al.* (215), Palsson *et al.* (288) and Papin *et al.* (289). Combined with linear programming techniques, flux distributions leading to optimization of; for instance, yields of product formation – such as biomass – or growth rate, can be predicted with flux balance analysis and compared to experimentally determined flux distributions (*e.g.* 31, 110, 394, 395).

2.6. Control and regulation

Systemic properties of cellular networks are determined by: (i) the organization of the network (spatial and interactional), (ii) the kinetic (*e.g.* K_M, V_{MAX}) and the physical properties (*e.g.* diffusion coefficients) of the components, and (iii) the environment (*e.g.* temperature, nutrient sources, or product sinks). Such constant aspects of cellular networks are generally referred to as parameters. Analysis of control of cellular networks involves the determination of the effects of changes in particular parameters on cellular functioning. Bifurcation analysis focuses on the behavior of dynamical systems as functions of parameters (128). Analysis of regulation determines how the system sustains or changes particular functions in response to changes in the environment of the system (147, 148, 182, 183). An analysis of control could address the effects of diffusion coefficients on fluxes in biochemical networks (291) or the effects of kinases and phosphatases on signaling (154). An analysis of regulation could

investigate how the cell maintains the ATP/ADP homeostasis in spite of continuous changes in its environment. Below we will see that the notions of control and of regulation are related.

Individual processes, such as a rate of an enzyme-catalyzed conversion or a molecular binding event, *control* systemic properties: when the rates of a process is changed – by changing a parameter that characterizes this process – systemic properties are affected quantitatively (in some cases even qualitatively, *e.g.* a change from a steady state to an oscillatory regime, *i.e.* a Hopf bifurcation (128, 145)). The rates of processes, and, therefore, also their control coefficients, depend nonlinearly on the state of the cellular network. In other words, the determination of the control properties of a cellular network necessitates the appreciation of the system as a whole (*in silico* or *in vivo*) to determine the state (*i.e.* $\mathbf{s}(t,\mathbf{p})$) and, subsequently, its responses to changes in parameters, such as levels of nutrients, growth factors, or hormones. Here we will focus on the control of cellular networks by their constituent processes. Emphasis on regulation of cellular networks can be found in Hofmeyr *et al.* (147, 148, 152) and Kahn & Westerhoff (182).

Metabolic control analysis (MCA) is a theoretical framework that is well suited to analyze control and regulation of cellular networks (102). It was pioneered by Heinrich & Rapoport (134) and Kacser & Burns (179) in the seventies for steady-state metabolic networks with early experimental application by Flint *et al.* (108, 109) and Groen *et al.* (127). MCA relates changes in systemic properties quantified by response or control coefficients to properties of enzymes in terms of elasticity coefficients. Also the control properties and elasticity coefficients of enzymes of a particular cellular network depend nonlinearly on the state of the cellular network. MCA has later been extended to include: control analysis of branched and cyclic pathways with and without moiety conservation (103, 104, 150, 180, 199, 200, 309, 333, 335, 337, 413), concentration control theory (410), energy coupling (416), control of generalized variables (343), control of transition times (85, 255), systems involving quasi-equilibrium reactions and time-scale separation (83, 201); oscillatory systems (159, 160, 191, 192), signaling networks (46, 194), channeling (189), intra-enzymatic processes (203), hierarchical networks with gene expression, signaling, and metabolism (46, 151, 168, 181, 360, 414), modular biochemical networks (347), reaction-diffusion systems (42, 291), and transient trajectories (1, 136, 161). It has been applied frequently to the experimental analysis of biochemical networks (*e.g.* 2, 3, 127).

2.7 Transient-state and steady-state control analysis

The control properties of the processes constituting the cellular network can be readily obtained from the differential equations (and the initial conditions) that govern the dynamics of the cellular network (Eq 2) by implicit differentiation with respect to the parameter vector that is perturbed (1, 136, 161):

$$\frac{d}{dt}\frac{\partial \mathbf{x}}{\partial \mathbf{p}}(t,\mathbf{p}) = \left(\mathbf{N}^0\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\right)\frac{\partial \mathbf{x}}{\partial \mathbf{p}}(t,\mathbf{p}) + \mathbf{N}^0\frac{\partial \mathbf{v}}{\partial \mathbf{p}} \qquad (8)$$

(Note that it is assumed that the reaction stoichiometries are not dependent on parameters but remain fixed.) The matrix of unscaled elasticity coefficients is defined as $\partial \mathbf{v}/\partial \mathbf{s} = \left[\partial \mathbf{v}/\partial \mathbf{x} \quad \partial \mathbf{v}/\partial \mathbf{x}^D\right] \equiv \tilde{\boldsymbol{\varepsilon}}$ and the Jacobian matrix as $\mathbf{N}^0\tilde{\boldsymbol{\varepsilon}}\mathbf{L}$. The values of the entries of both these matrices depend nonlinearly on the state of the cellular network.

Eq. 8 describes a system of linear ordinary differential equations with either constant or variable coefficients depending on the state of reference at which the parameter perturbation is applied. If the reference state of the cellular network at which the parameter perturbation is applied is: (i) a trajectory, the entries of the Jacobian matrix are time dependent and Eq. 8 has to be evaluated explicitly together with Eq. 2a and 2b (161) or (ii) an asymptotically stable steady state, the entries of the Jacobian matrix are time independent, and Eq. 8 can be evaluated on its own (136, 161). The initial conditions of Eq. 8 depend on whether kinetic parameters or initial conditions of Eq. 2 are perturbed. If an initial condition is perturbed, for instance of x_i, *i.e.* $dp=dx_i(0,\mathbf{p})$, the initial condition of the perturbation, $dx_i/dp(0,\mathbf{p})$, equals 1 and all others zero. If a kinetic parameter is perturbed all initial conditions of the dynamic description of the perturbed system (Eq. 8) are zero (161). Here only perturbations of kinetic parameters will be considered.

If the reference state is an asymptotically stable steady state, then the Jacobian matrix evaluated at the steady state of reference describes the dynamics of the system in the vicinity of the steady state (*e.g.* 376, 386). (More can be found on the analysis of stability of biochemical networks in Stucki (376), Tyson & Othmer (386), Higgins (145), Westerhoff & Van Dam (416), and Heinrich (133, 137).) This means that Eq. 8 can be evaluated independently of Eq. 2. In this case an analytical solution for this equation can be found (*i.e.* with $d\mathbf{x}/d\mathbf{p}(0)=0$) (136):

$$\frac{d\mathbf{x}}{d\mathbf{p}}(t,\mathbf{p}) = \left(\exp\left(\mathbf{N}^0\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\cdot t\right) - \mathbf{I}\right)\cdot\left(\mathbf{N}^0\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\right)^{-1}\cdot\mathbf{N}^0\cdot\frac{\partial \mathbf{v}}{\partial \mathbf{p}} \equiv \tilde{\mathbf{C}}^x_v(t)\cdot\frac{\partial \mathbf{v}}{\partial \mathbf{p}} \qquad (9)$$

This equation gives the concentration control coefficients in the network at each moment in time. Those coefficients are defined as:

$$\tilde{C}^{x_i}_{v_j}(t) = \frac{\dfrac{x_i(t,v_j+dv_j)-x_i(t,v_j)}{x_i(t,v_j)}}{\dfrac{v_j+dv_j-v_j}{v_j}}$$

Where $x_i(t,v_j)$ is the concentration of x_i in the unperturbed (reference) system, which remains in the reference steady state, *i.e.* $x_i(t,v_j)$ does not change over time. The concentration of x_i in the perturbed system, *i.e.* denoted by $x_i(t,v_j+dv_j)$, does depend on time and for long enough times will become constant: it will settle to a new steady state value (due to the assumed asymptotic stability of the reference steady state). The value for this concentration control coefficient in the new steady state equals the value of the classical control coefficient (defined in 410). This is illustrated in the following example. We consider the following network. $\longleftrightarrow^{1} X \longleftrightarrow^{2}$. The rate equations for the two enzymes are given by:

$$v_1 = \frac{1}{1+\dfrac{x}{0.5}}; \quad v_2 = \frac{x}{1+x}$$

The steady state concentration of x, *i.e.* x_{ss}, equals 0.7071 and the steady state flux v_{1ss} equals 0.414. At this steady state the unscaled elasticities equal $\tilde{\varepsilon}^{v1}_x = -0.343$, and $\tilde{\varepsilon}^{v2}_x = 0.343$ (it is a coincidence that these numbers are equal). The transient behavior of the scaled concentration control coefficient of the first enzyme (Eq. 9) is given by:

$$C^X_{v1}(t) = \frac{v_{1ss}}{x_{ss}}\frac{\left(e^{(\tilde{\varepsilon}^{v1}_x-\tilde{\varepsilon}^{v2}_x)vt}-1\right)}{\tilde{\varepsilon}^{v1}_x-\tilde{\varepsilon}^{v2}_x}$$

After a long enough period of time, this control coefficient equals the classical concentration control coefficient (1, 161, 410):

$$C^X_{v1}(\infty) = \frac{v_{1ss}}{x_{ss}}\frac{-1}{\tilde{\varepsilon}^{v1}_x-\tilde{\varepsilon}^{v2}_x}$$

Figure 5 displays the transient behavior of this control control coefficient and the dynamics of the perturbed network. Using the values of the steady state flux, steady state concentrations of x, and the elasticities of the reference steady state the aforementioned control coefficients equals, 0.85. (Using Figure 5A we can calculate also this control coefficient, *i.e.* ((0.7077-0.7071)/0.7071)/0.001=0.85.)

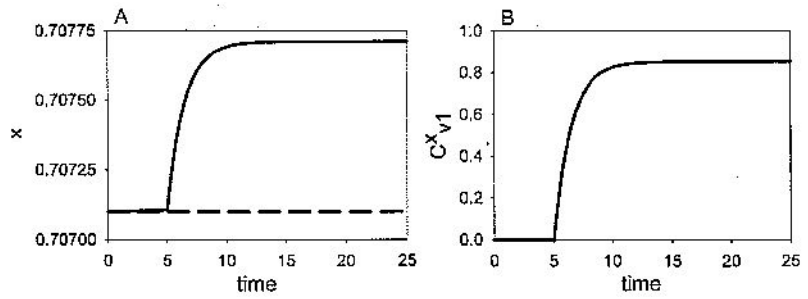

Figure 5.

Transient behavior of the two-enzyme and one metabolite pathway discussed in the text. A. the transient behavior of the network when the first enzyme is increased by factor of 1.001 in activity. The dashed line gives the steady state of reference at which the perturbation was applied. B. the transient behavior of the scaled concentration control coefficient of the first rate on the concentration of x.

The value of the (unscaled) flux control coefficients in the matrix $\tilde{\mathbf{C}}^J_v(t)$ at time t can be obtained from the implicit differentiation of $\mathbf{J} = \mathbf{v}(\mathbf{x}(t,\mathbf{p}),\mathbf{p})$ with respect to $\mathbf{p}$ (136):

$$\frac{d\mathbf{J}}{d\mathbf{p}}(t,\mathbf{p}) = \tilde{\boldsymbol{\varepsilon}}\mathbf{L}\cdot\frac{d\mathbf{x}}{d\mathbf{p}}(t,\mathbf{p}) + \frac{\partial \mathbf{v}}{\partial \mathbf{p}} = \left(\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\cdot\tilde{\mathbf{C}}^x_v(t) + \mathbf{I}\right)\frac{\partial \mathbf{v}}{\partial \mathbf{p}} = \tilde{\mathbf{C}}^J_v(t)\frac{\partial \mathbf{v}}{\partial \mathbf{p}} \qquad (10)$$

Multiplication of the time dependent (unscaled) concentration control matrix $\tilde{\mathbf{C}}^x_v(t)$ with the elasticity matrix ($\tilde{\boldsymbol{\varepsilon}}\mathbf{L}$) results in the connectivity theorem that applies during the relaxation of the perturbed system to the new steady state:

$$\tilde{\mathbf{C}}^x_v(t)\tilde{\boldsymbol{\varepsilon}}\mathbf{L} = \left(\exp\left(\mathbf{N}^0\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\cdot t\right) - \mathbf{I}\right) \qquad (11)$$

The connectivity theorem for (unscaled) flux control matrix has the following form:

$$\tilde{\mathbf{C}}^J_v(t)\tilde{\boldsymbol{\varepsilon}}\mathbf{L} = \tilde{\boldsymbol{\varepsilon}}\mathbf{L}\cdot\exp\left(\mathbf{N}^0\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\cdot t\right) \qquad (12)$$

The connectivity relationships relate the systemic control coefficients to the elasticity coefficients, which depend on the kinetic properties of the enzymes and the state of the network.

The summation theorems of the control coefficients are defined by the multiplication of the control matrices with the kernel of the stoichiometry matrix. For the concentration control coefficients summation theorem we obtain:

$$\tilde{\mathbf{C}}^x_v(t)\mathbf{K} = \mathbf{0} \qquad (13)$$

The summation theorem for the flux control coefficients is given by:

$$\tilde{\mathbf{C}}^J_v(t)\mathbf{K} = \mathbf{K} \qquad (14)$$

These two equations show that a simultaneous change in the rates of the processes that together form an independent flux mode for the system to reach a steady state (*i.e.* those processes that are combined in one column of **K**); (i) does not affect the values for **x** during the relaxation (Eq. 13) and (ii) the rates will change according to the entries in **K** during the relaxation (Eq. 14). These can be considered laws that restrict the total magnitude of the control exerted by processes on the concentrations and the rates within the system. They are generalization of the steady state flux and concentrations summation theorems.

The flux control coefficient matrix can be decomposed into two submatrices; that is, one containing the control coefficients on the independent and the other the control coefficients on the dependent fluxes. The connectivity theorem changes into (v^I stands for the rates of the independent reactions in the network):

$$\tilde{\mathbf{C}}^{J^I}_v(t)\tilde{\boldsymbol{\varepsilon}}\mathbf{L} = \frac{\partial \mathbf{v}^I}{\partial \mathbf{S}}\mathbf{L}\cdot\exp\left(\mathbf{N}^0\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\cdot t\right) \qquad (15)$$

The summation theorem of the flux control coefficients on the independent fluxes becomes:

$$\tilde{\mathbf{C}}^{J^I}_v(t)\mathbf{K} = \mathbf{I} \qquad (16)$$

Eqs. 13 to 16 can be compacted into one matrix equation that describes the control on the independent concentrations and independent flux in the network (136):

$$\begin{bmatrix}\tilde{\mathbf{C}}^{J^I}_v(t)\\ \tilde{\mathbf{C}}^x_v(t)\end{bmatrix}\begin{bmatrix}\mathbf{K} & -\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & -\frac{\partial \mathbf{v}^I}{\partial \mathbf{S}}\mathbf{L}\cdot\exp\left(\mathbf{N}^0\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\cdot t\right)\\ \mathbf{0} & \left(-\exp\left(\mathbf{N}^0\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\cdot t\right)+\mathbf{I}\right)\end{bmatrix} \qquad (17)$$

Frequently control coefficients and elasticities are used in a scaled format. Scaled control coefficients quantify the fractional change in fluxes or concentrations upon a fractional change in the activity of an enzyme. Scaled elasticity coefficients quantify the fractional change in the rate of an enzyme upon a fractional change in the concentration of a substrate, of a product or of an effector of that enzyme. To obtain scaled coefficients the matrices have to be changed in the following way:

$$\boldsymbol{\kappa} = Dg(\mathbf{J})^{-1}\cdot\mathbf{K}\cdot Dg(\mathbf{J}^I)$$

$$\boldsymbol{\varepsilon}\mathbf{L} = Dg(\mathbf{J})^{-1}\cdot\tilde{\boldsymbol{\varepsilon}}\mathbf{L}\cdot Dg(\mathbf{x})$$

$$\mathbf{C}^x_v = Dg(\mathbf{x})^{-1}\cdot\tilde{\mathbf{C}}^x_v\cdot Dg(\mathbf{J})$$

$$\mathbf{C}^J_v = Dg(\mathbf{J})^{-1}\cdot\tilde{\mathbf{C}}^J_v\cdot Dg(\mathbf{J})$$

We now obtain the following matrix equation:

$$\begin{bmatrix} \mathbf{C}_v^{J'}(t) \\ \mathbf{C}_v^{x}(t) \end{bmatrix}\begin{bmatrix} \boldsymbol{\kappa} & -\varepsilon\mathbf{L} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -Dg(\mathbf{J}')^{-1}\dfrac{\partial \mathbf{v}'}{\partial \mathbf{S}}\mathbf{L}\cdot\exp\left(\mathbf{N}^0\tilde{\varepsilon}\mathbf{L}\cdot t\right)Dg(\mathbf{x}) \\ \mathbf{0} & Dg(\mathbf{x})^{-1}\left(-\exp\left(\mathbf{N}^0\tilde{\varepsilon}\mathbf{L}\cdot t\right)+\mathbf{I}\right)Dg(\mathbf{x}) \end{bmatrix} \quad (18)$$

We have assumed the steady state to be asymptotically stable which means the real parts of all the eigenvalues of the Jacobian matrix are smaller than zero and that the matrix exponent $\exp\left(\mathbf{N}^0\cdot\tilde{\varepsilon}\mathbf{L}\cdot t\right)$ converges to a zero matrix for t→∞. Eq. 18 then simplifies into (204, 412):

$$\begin{bmatrix} \mathbf{C}_v^{J'} \\ \mathbf{C}_v^{x} \end{bmatrix}\begin{bmatrix} \boldsymbol{\kappa} & -\varepsilon\mathbf{L} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (19)$$

This equation is sometimes abbreviated by:

$$\mathbf{C}\cdot\mathbf{E} = \mathbf{I} \quad (20)$$

Interestingly the following relationship holds:

$$\mathbf{C} = \mathbf{E}^{-1} \qquad \left(\mathbf{E} = \mathbf{C}^{-1}\right) \quad (21)$$

Eq. 20 indicates that the control properties of system described by Eq. 2 can be obtained from its structural properties ($\mathbf{N}^0$, **K** and **L**) and kinetic properties ($\tilde{\varepsilon}$) by metabolic control analysis. This is an important result, since it relates changes in the systemic properties of biochemical network – the '*physiology*' – to changes in the activity of its components – the '*biochemistry*' – and the interaction structure – the '*organization*' – of the network.

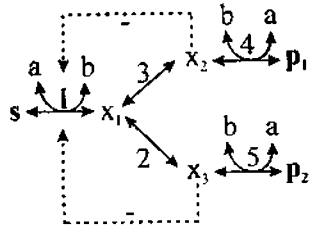

Figure 6.

A biochemical network composed of five reactions (v_1..v_5), five variable intermediates (x_1..x_3, a and b) and three boundary metabolites (S, P_1, and P_2) that are kept constant (indicated in bold). The first rate is inhibited by the intermediates x_2 and x_3 (denoted by a dashed arrow).

2.8 Example of the analysis of a biochemical reaction network

This section will illustrate the analysis of a kinetic model with the methods that have been outlined the previous subsections. The simple biochemical network displayed in Figure 6 will be analyzed. It consists of five reactions, five variable metabolites, and three fixed

metabolites.

2.8.1 Structural analysis

The network displayed in Figure 6 has the following stoichiometric matrix N:

$$\begin{pmatrix} & v1 & v2 & v3 & v4 & v5 \\ x1 & 1 & -1 & -1 & 0 & 0 \\ x2 & 0 & 0 & 1 & -1 & 0 \\ a & -1 & 0 & 0 & 1 & 1 \\ x3 & 0 & 1 & 0 & 0 & -1 \\ b & 1 & 0 & 0 & -1 & -1 \end{pmatrix}$$

The network contains three independent intermediates, which is reflected by the number of columns of the Link matrix **L**:

$$\begin{pmatrix} & x1 & x2 & a \\ x1 & 1 & 0 & 0 \\ x2 & 0 & 1 & 0 \\ a & 0 & 0 & 1 \\ x3 & -1 & -1 & -1 \\ b & 0 & 0 & -1 \end{pmatrix}$$

Here x_1, x_2, and a have been chosen as the independent intermediates. The moiety conservation relationships are given by the relationship:

$$\mathbf{x}^D - \mathbf{L}^t\cdot\mathbf{x}^I = \begin{pmatrix} x_3 \\ b \end{pmatrix} - \begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ a \end{pmatrix} = \begin{pmatrix} a + x_1 + x_2 + x_3 \\ a + b \end{pmatrix}$$

There are therefore two pools of metabolites, the obvious a+b and the less intuitive $a+x_1+x_2+x_3$, that remain constant during any behavior of the biochemical network. The reduced stoichiometric matrix $\mathbf{N}^0$, which contains only the stoichiometric coefficients of the independent intermediates, is given by:

$$\begin{pmatrix} & v1 & v2 & v3 & v4 & v5 \\ x1 & 1 & -1 & -1 & 0 & 0 \\ x2 & 0 & 0 & 1 & -1 & 0 \\ a & -1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

When the network is in steady state the number of independent fluxes are given by the number of columns of Kernel matrix **K**:

$$\begin{pmatrix} & J_2 & J_3 \\ J_2 & 1 & 0 \\ J_3 & 0 & 1 \\ J_1 & 1 & 1 \\ J_4 & 0 & 1 \\ J_5 & 1 & 0 \end{pmatrix}$$

The independent fluxes in steady state correspond to J_2 and J_3. Note that J_4 and J_5 could have been chosen as well. Each column of the kernel matrix corresponds to a pathway in the network that can independently attain a steady state. The first column corresponds to the overall conversion of S into P_2 and the second column corresponds to the conversion of S into P_1. (Had we added the conversion of ATP into ADP to reaction 3 and would we have kept the concentration of ATP and ADP fixed we would have obtained for the second column the overall conversion S+ATP→ADP+P_1. Such kind of analysis allows for the discrimination of alternative pathways on basis of energy expenditure for instance.)

The overall reaction of the two independent flux modes can be obtained from:

$$\mathbf{K}^T \cdot \mathbf{R} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} -x_1 + x_3 \\ -x_1 + x_2 \\ -a + b - s + x_1 \\ a - b + p_1 - x_2 \\ a - b + p_2 - x_3 \end{pmatrix} = \begin{pmatrix} p_2 - s \\ p_1 - s \end{pmatrix}$$

Where **R** is defined as the vector that contains the j-th reaction as its j-th entry and $\sum_i n_{ij} \cdot s_i + \sum_i n_{ij} \cdot p_i$ with n_{ij} as the (i,j)-th entry of N and with s_i and p_i as the substrates and products of the j-th reaction. The chemical potential of the net conversions in a particular steady state can be obtained by (302):

$$\mathbf{K}^T \cdot \Delta\boldsymbol{\mu} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \Delta\mu_2 \\ \Delta\mu_3 \\ \Delta\mu_1 \\ \Delta\mu_4 \\ \Delta\mu_5 \end{pmatrix} = \begin{pmatrix} \Delta\mu_1 + \Delta\mu_2 + \Delta\mu_5 \\ \Delta\mu_1 + \Delta\mu_3 + \Delta\mu_4 \end{pmatrix} = RT \begin{pmatrix} \ln \dfrac{p_2}{s \cdot K_{EQ,1} \cdot K_{EQ,2} \cdot K_{EQ,5}} \\ \ln \dfrac{p_1}{s \cdot K_{EQ,1} \cdot K_{EQ,3} \cdot K_{EQ,4}} \end{pmatrix}$$

On the basis of the signs of the entries in the last matrix it can be decided in which direction the flux runs at steady state. If for instance the first entry is negative then mass will flow from s to p_2. If entries in this vector are evaluated as zero in a particular steady state, then the corresponding reactions form a subsystem at equilibrium.

Suppose that enzyme 5 is irreversible. The extreme pathway matrix is then identical

to the kernel matrix. But the elementary mode matrix contains one extra column:

$$\begin{pmatrix} J_2 & 1 & 0 & 1 \\ J_3 & 0 & 1 & -1 \\ J_1 & 1 & 1 & 0 \\ J_4 & 0 & 1 & -1 \\ J_5 & 1 & 0 & 1 \end{pmatrix}$$

The extreme pathway matrix contains the minimal number of steady state flux modes with all reaction operating in their preferred thermodynamic direction. The elementary mode matrix contains all those flux modes that can independently generate a steady state, with all rates operating in their thermodynamically preferred direction, and with all flux modes as non-decomposable. Non-decomposable means that the removal of one reaction from the elementary would yield a flux mode that cannot reach a steady state on its own. (Note that the third elementary mode can be obtained from a linear combination of the extreme pathway matrix.) The overall reactions and their Gibbs free energy difference at steady state can be obtained for the elementary flux modes and extreme pathways in a similar manner as has been shown above in terms of the kernel matrix.

2.8.2. Control analysis

The topology of mass flow regulation of the network is described with the unscaled elasticity matrix $\tilde{\boldsymbol{\varepsilon}}$, given by:

$$\begin{pmatrix} & x1 & x2 & a & x3 & b \\ v1 & \varepsilon_{x1}^{v1} & \varepsilon_{x2}^{v1} & \varepsilon_{a}^{v1} & \varepsilon_{x3}^{v1} & \varepsilon_{b}^{v1} \\ v2 & \varepsilon_{x1}^{v2} & 0 & 0 & \varepsilon_{x3}^{v2} & 0 \\ v3 & \varepsilon_{x1}^{v3} & \varepsilon_{x2}^{v3} & 0 & 0 & 0 \\ v4 & 0 & \varepsilon_{x2}^{v4} & \varepsilon_{a}^{v4} & 0 & \varepsilon_{b}^{v4} \\ v5 & 0 & 0 & \varepsilon_{a}^{v5} & \varepsilon_{x3}^{v5} & \varepsilon_{b}^{v5} \end{pmatrix}$$

The Jacobian matrix $\mathbf{N}^0\tilde{\boldsymbol{\varepsilon}}\mathbf{L}$ of this particular network has the following form:

$$\begin{pmatrix} \varepsilon_{x1}^{v1} - \varepsilon_{x1}^{v2} - \varepsilon_{x1}^{v3} - \varepsilon_{x3}^{v1} + \varepsilon_{x3}^{v2} & \varepsilon_{x2}^{v1} - \varepsilon_{x2}^{v3} - \varepsilon_{x3}^{v1} + \varepsilon_{x3}^{v2} & \varepsilon_{a}^{v1} - \varepsilon_{b}^{v1} - \varepsilon_{x3}^{v1} + \varepsilon_{x3}^{v2} \\ \varepsilon_{x1}^{v3} & \varepsilon_{x3}^{v2} - \varepsilon_{x2}^{v4} & -\varepsilon_{a}^{v4} + \varepsilon_{b}^{v4} \\ -\varepsilon_{x1}^{v1} + \varepsilon_{x1}^{v1} - \varepsilon_{x3}^{v5} & -\varepsilon_{x2}^{v1} + \varepsilon_{x2}^{v4} + \varepsilon_{x3}^{v1} - \varepsilon_{x3}^{v5} & -\varepsilon_{a}^{v1} + \varepsilon_{a}^{v4} + \varepsilon_{a}^{v5} + \varepsilon_{b}^{v1} - \varepsilon_{b}^{v4} + \varepsilon_{x3}^{v1} - \varepsilon_{x3}^{v5} \end{pmatrix}$$

The Jacobian matrix plays a dominant role in the analysis of dynamical systems. It can be used for, e.g. local stability analysis by evaluating the eigenvalues of this matrix (128), metabolic control analysis (see previous section), and the determination of the response of the system to step and sinusoidal changes in parameters with frequency analysis (160).

The E matrix of MCA that equals the augmented matrix composed out of the scaled kernel matrix and the scaled elasticity matrix, *i.e.* $(\kappa \quad -\varepsilon L)$, is now given by:

$$\begin{pmatrix}
 & J_2 & J_3 & x_1 & x_2 & a \\
J_2 & 1 & 0 & -\varepsilon_{x1}^{v1}+\frac{x_1}{x_3}\varepsilon_{x3}^{v1} & -\varepsilon_{x2}^{v1}+\frac{x_2}{x_3}\varepsilon_{x3}^{v1} & -\varepsilon_{a}^{v1}+\frac{a}{b}\varepsilon_{b}^{v}+\frac{a}{x3}\varepsilon_{x3}^{v1} \\
J_3 & 0 & 1 & -\varepsilon_{x1}^{v2}+\frac{x_1}{x_3}\varepsilon_{x3}^{v2} & \frac{x_2}{x_3}\varepsilon_{x3}^{v2}, & \frac{a}{x_3}\varepsilon_{x3}^{v2} \\
J_1 & \frac{J_2}{J_1} & \frac{J_3}{J_1} & -\varepsilon_{x1}^{v3} & -\varepsilon_{x2}^{v3} & 0 \\
J_4 & 0 & \frac{J_3}{J_4} & 0 & -\varepsilon_{x2}^{v4} & -\varepsilon_{a}^{v4}+\frac{a}{b}\varepsilon_{b}^{v4} \\
J_5 & \frac{J_2}{J_5} & 0 & \frac{x_1}{x_3}\varepsilon_{x3}^{v5} & \frac{x_2}{x_3}\varepsilon_{x3}^{v5} & -\varepsilon_{a}^{v5}+\frac{a}{b}\varepsilon_{b}^{v5}+\frac{a}{x_3}\varepsilon_{x3}^{v5}
\end{pmatrix}$$

(Note that in this equation the elasticity matrix has been supplied with a different ordering of the rows, identical the ordering in the K matrix.) The inverse of this matrix gives the control matrix, which contains the control coefficients of all rates on the independent fluxes and concentrations in the network as a matrix:

$$\begin{pmatrix}
C_{v2}^{J2} & C_{v3}^{J2} & C_{v1}^{J2} & C_{v4}^{J2} & C_{v5}^{J2} \\
C_{v2}^{J3} & C_{v3}^{J3} & C_{v1}^{J3} & C_{v4}^{J3} & C_{v5}^{J3} \\
C_{v2}^{x1} & C_{v3}^{x1} & C_{v1}^{x1} & C_{v4}^{x1} & C_{v5}^{x1} \\
C_{v2}^{x2} & C_{v3}^{x2} & C_{v1}^{x2} & C_{v4}^{x2} & C_{v5}^{x2} \\
C_{v2}^{a} & C_{v3}^{a} & C_{v1}^{a} & C_{v4}^{a} & C_{v5}^{a}
\end{pmatrix}$$

2.8.3 Dynamic analysis

The reactions were described with the following rate equations:

$$v1=\frac{V_1\cdot\frac{a}{K_{1a}}\cdot\frac{S}{K_{1s}}\cdot\left(1-\frac{b\cdot x1}{a\cdot S\cdot K_{eq}}\right)}{\left(1+\left(\frac{x2}{K_{1x2}}\right)^5+\left(\frac{x3}{K_{1x3}}\right)^5\right)\left(1+\frac{a}{K_{1a}}+\frac{b}{K_{1b}}\right)\left(1+\frac{S}{K_{1s}}+\frac{x1}{K_{1x1}}\right)},$$

$$v2=\frac{V_2\cdot\frac{x1}{K_{2x1}}\cdot\left(1-\frac{x3}{x1\cdot K_{eq}}\right)}{1+\frac{x1}{K_{2x1}}+\frac{x3}{K_{2x3}}},\quad v3=\frac{V_3\cdot\frac{x1}{K_{3x1}}\cdot\left(1-\frac{x2}{x1\cdot K_{eq}}\right)}{1+\frac{x1}{K_{3x1}}+\frac{x2}{K_{3x2}}},$$

$$v4=\frac{V_4\cdot\frac{b}{K_{4b}}\cdot\frac{x_2}{K_{4x2}}\cdot\left(1-\frac{a\cdot P}{b\cdot x2\cdot K_{eq}}\right)}{\left(1+\frac{a}{K_{4a}}+\frac{b}{K_{4b}}\right)\left(1+\frac{P}{K_{4P}}+\frac{x2}{K_{4x2}}\right)},\quad v5=\frac{V_5\cdot\frac{b}{K_{5b}}\cdot\frac{x_3}{K_{5x3}}\cdot\left(1-\frac{a\cdot P}{b\cdot x3\cdot K_{eq}}\right)}{\left(1+\frac{a}{K_{5a}}+\frac{b}{K_{5b}}\right)\left(1+\frac{P}{K_{5p}}+\frac{x3}{K_{5x3}}\right)}$$

(Note that reaction 5 has been chosen reversible again.) The transient behavior of the network is displayed in Figure 7.

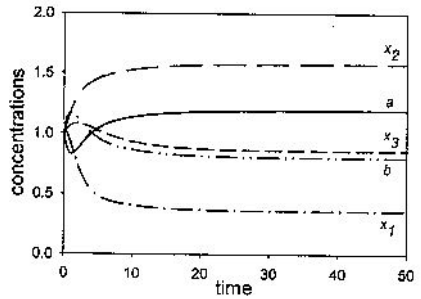

Figure 7.

Transient behavior of the network displayed in Figure 6. The kinetic constants and boundary conditions were chosen as: K=1 (all K's), S=5, P=1, Keq=100, V_1=10, V_2=1, V_3=1, V_4=1, V_5=2. The following initial conditions were chosen: a[0]=1, b[0]=1, x_1[0]=1, x_2[0]=1, x_3[0]=1.

In steady state, the following concentrations were calculated:

x_1=0.360, x_2=1.58, a=1.20, x_3=0.86, b=0.80

Note that indeed the pool concentrations are conform the moieties that were found given the

initial conditions of the network. The fluxes were calculated as:

$J_1=0.28$, $J_2=0.16$, $J_3=0.12$, $J_4=0.12$, $J_5=0.16$

We see that the following relationships hold; $J_1=J_2+J_3$, $J_4=J_3$, and $J_5=J_2$. This confirms that there exist only two independent fluxes in the system (*i.e.* three relations exist for five fluxes) in accordance with the kernel **K** that was found above for this network. Also this flux distribution is a linear combination of the flux modes ($(\mathbf{J}^{I},\mathbf{J}^{D})^{T}=\mathbf{K}\mathbf{J}^{I}$), the elementary flux modes (with nonnegative coefficients for the linear combination that sum to unity (344)), and the extreme pathways (with coefficients for the linear combination that form the so-called α-spectrum); for a discussion see Papin *et al.* (289). Note that in extreme pathway analysis and elementary mode analysis the coefficients used in the linear combination are not unique and that they are always nonnegative (289).

2.8.4 Steady-state control analysis

In the steady state, the control coefficients were calculated for the extent to which each process controlled any of the independent fluxes and concentrations (an (i,j)-entry in this matrix refers to the control on variable i by process j):

	$v2$	$v3$	$v1$	$v4$	$v5$
$J2$	0.81	−0.95	0.42	0.77	−0.06
$J3$	0.22	−0.41	0.43	1.22	−0.46
$x1$	0.10	−1.34	0.51	1.12	−0.39
$x2$	−0.21	0.29	0.06	−0.32	0.18
a	−0.22	0.38	−0.26	−0.27	0.37

The entire flux control matrix, $\mathbf{KC}_{v}^{J^{I}}$, is given by:

	$J2$	$J3$	$J1$	$J4$	$J5$
$J2$	0.81	−0.95	0.42	0.77	−0.06
$J3$	0.22	−0.41	0.43	1.22	−0.46
$J1$	0.56	−0.72	0.43	0.96	−0.23
$J4$	0.22	−0.41	0.43	1.22	−0.46
$J5$	0.81	−0.95	0.42	0.77	−0.06

The entire concentration control matrix, $\mathbf{LC}_{v}^{x^{I}}$ is given by:

	$v2$	$v3$	$v1$	$v4$	$v5$
$x1$	0.10	−1.34	0.51	1.12	−0.39
$x2$	−0.21	0.29	0.06	−0.32	0.18
a	−0.22	0.38	−0.26	−0.27	0.37
$x3$	0.66	−0.51	0.03	0.49	−0.68
b	0.33	−0.57	0.39	0.40	−0.56

One would expect enzymes to have positive control on their own rates. (Someone who is more inclined to think reductionistically would go as far as to think that each enzyme should control its own rate with a control coefficient of 1: not taking into account that also the other enzymes exert control and that the sum of control is bounded by the summation theorems of MCA). However, in this particular example we observe that enzymes 3 and 5 exert negative control on all fluxes of the system (including their own flux). In this example additional violations of rules of thumb occur. The following 'intuitions' are violated: (i) enzymes control their own rate positively and (ii) in branched systems the enzymes that are part of the same branch have positive control on each other's fluxes and negative control on the fluxes within other branches. The reasons for such violations of 'basic intuitions' often derive from the existence of feedback and feedforward loops, or from moiety conservation relationships. For instance, the model with the feedback regulation (*i.e.* the effects of x_3 and x_3 on v_1 removed) removed has the following flux and concentration control matrix:

	$v2$	$v3$	$v1$	$v4$	$v5$
$J2$	0.53	−0.25	0.13	0.20	0.40
$J3$	−0.15	0.31	0.14	0.63	0.07
$J1$	0.24	−0.01	0.13	0.39	0.26
$J4$	−0.15	0.31	0.14	0.63	0.07
$J5$	0.53	−0.25	0.13	0.20	0.40

	$v2$	$v3$	$v1$	$v4$	$v5$
$x1$	−0.34	−0.57	0.19	0.49	0.23
$x2$	−0.22	0.55	0.01	−0.56	0.22
a	0.18	−0.10	−0.82	0.46	0.29
$x3$	0.80	−0.39	−0.01	0.39	−0.80
b	−0.03	0.02	0.13	−0.07	−0.05

However, we still observe a violation of the second rule of thumb. For instance, rate 5 does not negatively control the flux through the other branch. The second rule of thumb is no longer violated if, in addition, the moiety-conservation relationships are removed, that is, by

removal of a and b from the reactions of the model. Then we find:

$$\begin{pmatrix} & v2 & v3 & v1 & v4 & v5 \\ J_2 & 0.85 & -0.02 & 0.10 & -0.02 & 0.10 \\ J_3 & -0.07 & 0.50 & 0.14 & 0.43 & -0.01 \\ x1 & -0.51 & -0.26 & 1.05 & -0.22 & -0.06 \\ x2 & -0.27 & 2.06 & 0.56 & -2.32 & -0.03 \\ x3 & 1.53 & -0.04 & 0.18 & -0.04 & -1.62 \end{pmatrix}$$

In summary, many rules of thumb do not consider the effects of feedback loops and moiety conservation relationships in their argumentation. Unfortunately, feedback loops and moiety conservation relationships are ubiquitous in cellular networks, most noticeably in signaling and metabolic networks. Therefore, mathematical models of cellular networks are essential for the evaluation of control and regulation if the networks considered are intricately organized structures with many feedback loops, branches, cycles, and moiety-conservation relationships.

2.8.5 Transient-state control analysis

A component of metabolic control analysis that has not been given much attention in the past is the control of transient phenomena (noticable exceptions have been 1, 136, 161, 191). This part of control analysis may become important in signal transduction where transient states, more so than steady states, are perceived as the functional states. Even though in metabolism, steady states are considered the functional states, transient control analysis also becomes important there to understand adaptive physiological changes. Figure 8 displays all the concentration control coefficients as function of time during the relaxation towards the steady state for the system displayed in Figure 7. The control coefficients were obtained from integration of the differential equations and simultaneous evaluation of Eq. 8. The control coefficients portray the difference between the dynamics of the unperturbed network and the network where one maximal rate has been increased by a differentially small amount. The networks had the same initial conditions – hence, the only difference is the perturbed parameter. Since, there are five reactions in the network, five maximal rates have been independently perturbed, *i.e.* not all at the same time.

The sum of the transient concentration control coefficients on x_j are a particular time t has been shown to equal the scaled rates of change of x_j at time t provided that time zero is chosen to represent the time point at which the perturbation of the rates were performed (154, 191, 291):

$$\sum_i C_{v_i}^{x_j}(t) = \frac{d \ln x_j}{d \ln t}(t)$$

Performing this summation at time 5 and 20 we obtain the results shown in Table 3. The control coefficients at time 20 are already quite similar to their steady-state values that were shown in the matrix above.

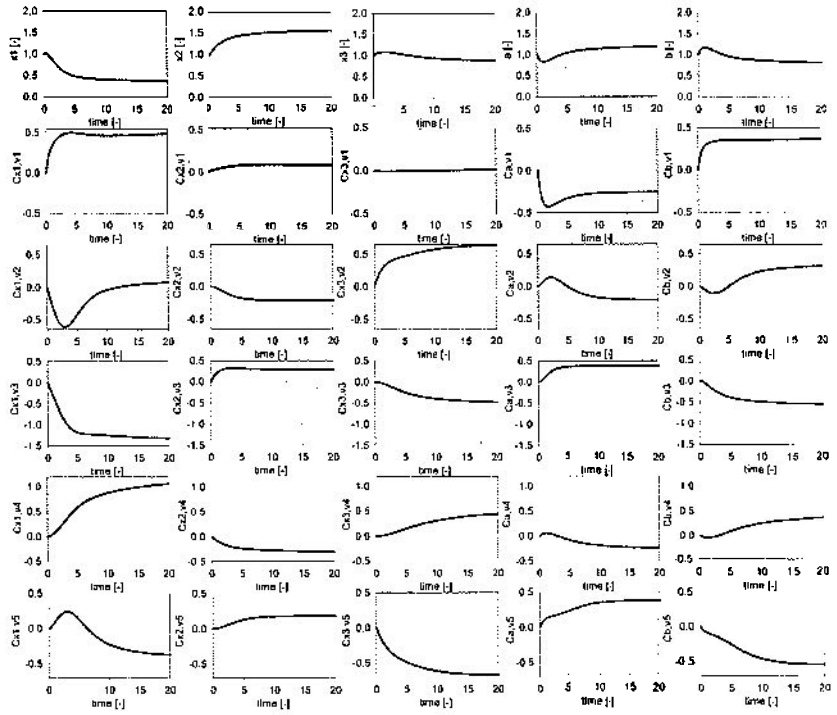

Figure 8.

Transient behavior of the state of the unperturbed network (first row of figures; and the evolution of the scaled concentration control coefficients (remaining figures).

Table 3. *Transient concentration control coefficients at two points in time and an illustration of the summation theorem.*

Control\time	5	20
C_1^{X1}	0.50	0.49
C_2^{X1}	-0.42	0.075
C_3^{X1}	-1.19	-1.32
C_4^{X1}	0.59	1.06
C_5^{X1}	0.13	-0.37
$\sum_i C_i^{X1}$	-0.39	-0.074
$\frac{d \ln X_1}{d \ln t}$	-0.39	-0.074

All the calculations within this section have been performed with a kinetic modeling package developed in *Mathematica*. This kinetic modeling package does all of the calculations displayed in this section (and a lot more) with a minimal input of the biochemical network. It can be obtained together with the example model shown above from the author. The full model including feedbacks and moiety conservation relationships can also be obtained from the author as a file to analyze with Gepasi, Jdesigner, or in SBML format. The model can be investigated online at the JWS online website (http://jjj.biochem.sun.ac.za) (more information can be obtained from the author).

2.9 The organization of biochemical reaction networks

2.9.1 Small-world organization

Recently, many studies have analyzed the (nearly) system-wide organization of various cellular networks, *e.g.* metabolism (excluding allosteric interactions) (7, 105, 173, 296, 308, 399), genetic circuitry (105, 173, 261, 296, 357, 399), and the protein interaction network (172, 293, 365, 432). All cases have in common that they consider the networks as (directed or undirected) graphs with the molecular species as nodes and the interactions/conversions as edges. In the analysis of metabolism these studies do not consider the presence of allosteric interactions and moiety conservation (for inclusion of moiety conservations and its consequences see (7)). These studies have shown that many networks have a common underlying scale-free organization (19, 105, 173, 296, 404) (for an exception see (7)). In

random networks[2] the degree distribution p_k, *i.e.* the probability of a node to have k edges: the fraction of nodes with k edges, is binomially distributed, which approaches a Poison distribution for large numbers of nodes (n) (at constant k), *i.e.* $p_k=(z^k e^{-z})/k!$ with z as the mean number of edges per node (4, 276). The only property of that makes networks scale free is the fact that their degree distribution is not a function of the number of nodes but only of k, *i.e.* $p_k \sim k^{-\alpha}$ with an exponent α which is characteristic for the network and independent of the number of nodes n (4, 276). The exponent for a metabolic network composed of 765 nodes and 3686 undirected edges was found to be 2.2 (173). In other words, these networks have only a few nodes with a high number of edges and many nodes with a small number of edges. As a consequence, they have short average path lengths between nodes (just like many random graphs). In addition, they proved to display a high level of clustering. The clustering coefficient C_i of node i with k_i edges is defined as the actual number of edges between its k_i neighbours divided by the maximal number of edges possible between its neighbours, which is given by $k_i(k_i-1)/2$ (404). The clustering coefficient of the entire network is given by $C=\frac{1}{n}\sum_i C_i$. For a random network the clustering coefficient is given by the probability for an edge p, *i.e.* $C=z/n$. For the small-world network of Barabasi & Albert (19) the clustering coefficient is found to be ~5 times higher than the random graph models (4). Scale-free networks with these two properties, *i.e.* scale-freeness and clustering, are frequently referred to as small-world networks. The metabolic network of *E. coli* and many other cellular networks, have been show to be small world (105, 173, 296, 399).

The success of understanding the behaviors of systems in terms of the behaviors of their components is not self-evident. There are many different decompositions of systems into parts possible and only some of them are sufficiently insightful to lead to understanding of how systems work (185, 424). Likewise, to understand how a computer program works, it is not always most insightful to look at its binary code. This illustrates that the search for functional organization – the organization that illustrates the roles of components – rather than merely the physical organization is not necessarily obvious and intuitive. Functional

[2] One way to construct random undirected networks is to connect each pair of nodes in an n node network with a probability p. The probability to obtain a graph with m edges is then given by $p^m(1-p)^{M-m}$ with $M=(n(n-1))/2$ as the maximum number of edges. The mean degree (number of edges per node) is given by $z=\sum_k k \cdot p_k = p(n-1)$.

organization is also not necessarily in terms of the physical or isolable constituents of a system but may be in terms of functional units that are more vague and dynamic of shape (420). So far it is unclear what the functional consequences are of scale-free organization of biological networks or whether it is merely a reflection of the process that led to them, *i.e.* of natural selection and of random mutation.

2.9.2 Modular organization & modular control analysis

Many studies have focused on the modular organization of cellular networks (3, 46, 112, 131, 151, 202, 238, 293, 307, 308, 319, 325, 347, 348, 357, 365, 381, 407). Frequently, the modules defined in these studies correspond to subsystems with components that have more interactions among themselves than with extramodular components (308). However, the underlying motivation for the search for modules is in most cases the hope to find functional modules or, even, motifs that perform a specific task within the cell (131, 387, 407, 426). Three functional motifs consisting of either 3 or 4 nodes were found in the genetic circuit of *E. coli* (357) on the basis of their high probability of occurrence as compared to their probability of occurrence in an ensemble of thousand random networks that had the same number of nodes as the genetic circuit of *E. coli* (n=424 and k=577) (for an alternative method see (163)). In particular, the feedforward motif (the other two are the single-input module, and dense-overlapping regulons) has interesting properties (242, 243). It remains to be seen, however, to what extent the functional properties of these recurrent subnetworks are invariant with respect to the strength of the interactions they have with other subnetworks in living cells. In other words, the robustness of motif function is an interesting property to investigate.

None of the structural methods based on graph theory consider the strength of the interactions and in most cases the kinetic parameters of processes that occur within motifs are unknown. Therefore, the usefulness of the identification of functional motifs on structural information alone remains to be shown.

The modular methods derived in control analysis have a more functional perspective. Modular control analysis studies the responses of cellular networks in terms of interactions between modules (407). It gives quantitative rules for the definition of modules within networks. The criteria for modular organization of steady-state cellular networks derived within control analysis on the basis of the so-called structural properties and interaction structures, will be discussed now. The structural properties of cellular networks are the kinetically invariant aspects of the network. Examples are the stoichiometric matrix **N**, its

kernel matrix **K**, and the link matrix **L**. The interaction structure of the network is reflected by the elasticity matrix $\boldsymbol{\varepsilon}$ or $\boldsymbol{\varepsilon L}$.

Control analysis makes use of the principle that a subsystem (obeying certain criteria that will be mentioned below) attains the same steady state when the entire system is allowed to relax to a new (global) steady state upon a parameter perturbation, as it does when all the extramodular intermediates it communicates with are changed to their steady-state values that correspond to the perturbed global state, and the subsystem is then allowed to relax. For subsystems that are directly affected by the perturbed parameter, the parameter also has to be perturbed to the value that corresponds to the perturbed state. Subsystems display this characteristic if they obey two conditions to be mentioned below (cf. 137, 347). These subsystems can be treated as modules within control analysis.

The first condition is that moiety-conservation relationships only occur within modules: molecules that are part of different modules should not be engaged in the same moiety-conservation relationship (137, 347). If a moiety-conservation relationship crosses modular boundaries, the subsystem attains a local steady state when its extramodular communicating intermediates and the parameter are perturbed, that is different from its steady state when the entire system is allowed to relax, because the moiety is distributed differently over its molecular 'carriers' in the two cases.

The second condition is that all intermediates that affect extramodular or bridging reactions through a regulatory interaction, *e.g.* inhibitory/activatory feedforward or feedback interactions, or, product or substrate interactions, should be treated as communicating intermediates, respectively. In other words, these intermediates should be considered explicitly. Interactions between a reaction and an effector metabolite that are part of the same module – intramodular interactions – may remain 'hidden' within modules. The latter possibility allows modules to be treated as 'black boxes': only the variables that affect processes outside of the module, *i.e.* the communicating intermediates that affect extramodular and bridging reactions, are considered in the analysis; the intermediates that are solely involved in intramodular processes and the intramodular processes themselves are not treated explicitly. Interestingly, the modules can also be considered as translucent boxes such that their modular properties can be expressed in terms of their internal and external interactions (46, 151, 181, 347)

The two conditions for modular analysis are reflected in the organization of the **L** matrix and the elasticity matrix $\boldsymbol{\varepsilon}$ (or $\boldsymbol{\varepsilon L}$) of the network. Subsystems that have separate moiety conservation relationships, such that no molecules of different subsystems occur in the

same moiety conservation relationship, are reflected by an L-matrix that can be block diagonalized. Therefore, cellular networks that can be modularized have a block-diagonalizable L matrix. The nonzero entries or blocks within the L matrix correspond to groups of molecules that can be chosen to form one module.

Scheme A (Figure 9) can be modularized into two modules on the basis of the blocks in its L matrix, *i.e.* one module comprising x_1, x_2, v_1 and v_2 and one comprising x_3, x_4, v_3, and v_4. Intermediate x_2 is the communicating intermediate in this system. There are no bridging reactions that transport mass between the subsystems. This is apparent both from the reaction scheme A and the block-diagonal K matrix. The block-diagonal character of the K matrix derives from the fact that in steady state the two modules will not share net mass flow (181).

Scheme B (Figure 9) is a more detailed version of scheme A. In scheme B the enzymes are treated not as a whole but as networks composed of substrate complexes and single free enzymes. Scheme B also has a block-diagonal K matrix indicating that there is no net mass transport between the two subsystems of the network that each contain the reactions associated with one of the blocks in the K matrix. (Note that this was not immediately obvious from the biochemical scheme.) However, the L matrix cannot be block-diagonalized indicating that none of the subsystems correspond to a true module. Interestingly, the shift in detail from scheme A to scheme B leads to a loss of effective modular functioning (cf. 151).

There are two types of modules that correspond to blocks within the L matrix. The first type is a level. These modules contain rates that form a block within the K matrix. Kahn & Westerhoff (181) initially took diagonal blocks in the stoichiometric matrix as conditions for level-modularity, but this was later proven to be a sufficient rather than a necessary condition by Heinrich and Schuster (137): they showed the block-diagonality of the K matrix to be the necessary condition. These kinds of modules do not exchange net mass flow with other modules in steady state. Level-type modular systems have interesting control properties as has been analyzed by (44, 46, 151, 181, 194, 197). The second type of module does not have a block-diagonal K matrix, but solely a block-diagonal L matrix. These kinds of modules have been analyzed by (35, 325, 347, 407, 416).

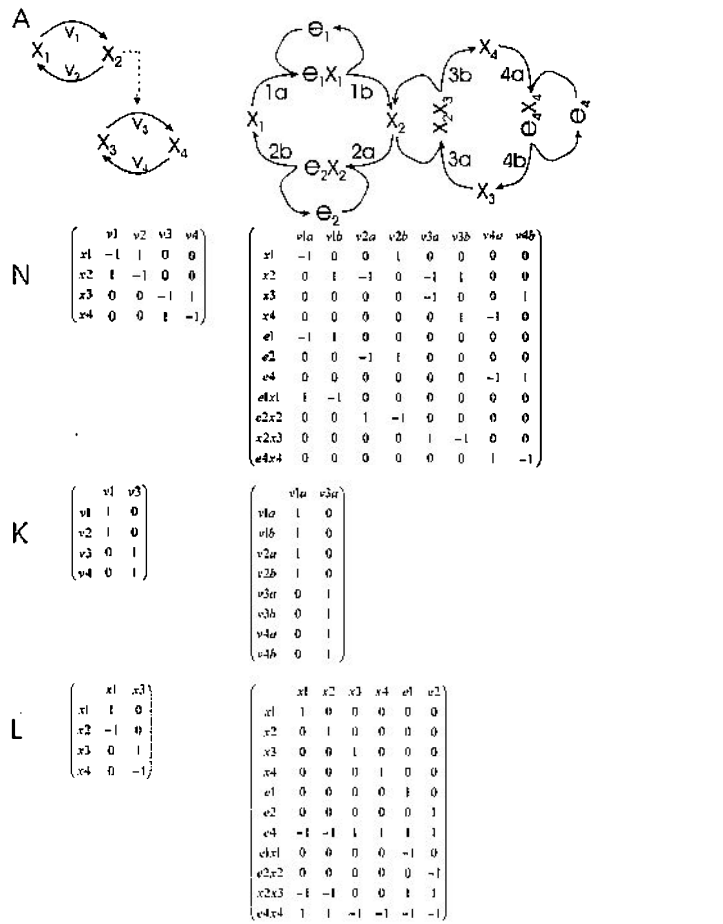

Figure 9.

Two tier cascade. The biochemical scheme depicted in A is depicted in more detail in B. All the matrices in Figure A are block diagonal whereas only the kernel can be block diagonalized in Figure B.

3. Ammonium assimilation by *Escherichia coli*

Ammonium is considered the preferred nitrogen source for *E. coli*, since all other nitrogen sources result in slower growth rates. Ammonium is transported across the plasma membrane via passive diffusion or via the transport protein AmtB (362, 364, 382). The mechanism by which AmtB transports ammonium, *i.e.* facilitated diffusion or active transport, is still under experimental investigation. It has been shown that the AmtB transporter is active only under

severe limitation of ammonium (66, 165). At $[NH_4^+]<5\mu M$ the AmtB transporter is active and above 50 μM it is inactivated by the binding of a protein called GlnK (66, 165, 390) (see below). The level of GlnK increases with the intracellular nitrogen status as monitored by glutamine. GlnK and AmtB are part of the same operon glnKAmtb (382). The activity of the ammonium transporter is regulated via an intricate mechanism by GlnK (66). The cytosolic processes associated with ammonium assimilation are displayed in Figure 10 excluding ammonium transport.

E. coli uses two mechanisms to assimilate ammonium. One is the GS/GOGAT system composed of glutamine synthetase (GS) and glutamate synthase (GOGAT) (315),

$$NH_4^+ + GLU + ATP \xleftrightarrow{GS} GLN + ADP + Pi$$

$$GLN + KG + NADPH \xleftrightarrow{GOGAT} 2\,GLU + NADP$$

(With GLN referring glutamine and KG to α-ketoglutarate.) The second mechanism involves glutamate dehydrogenase (GDH) (315),

$$NH_4^+ + KG + NADPH \xleftrightarrow{GDH} GLU + NADP$$

The net difference between the two mechanisms is the net hydrolysis of one molecule of ATP per molecule of ammonium assimilated by GS/GOGAT. This difference makes GS/GOGAT the energetically unfavorable alternative (138, 139). GS/GOGAT is preferentially used under ammonium limitation and GDH gradually takes over ammonium assimilation upon an increase in the ammonium concentration (315). GS has a higher affinity for ammonium than GDH does (254, 328). At high ammonium concentrations, GS is inactivated by reversible adenylylation of all of its twelve subunits (132, 208): the maximal rate of GS decreases in an inversely sigmoidal fashion (117). There are only minor cooperative effects among the twelve subunits of GS (84). The bifunctional enzyme that acts as the covalent modifier and demodifier is adenylyltransferase (ATase) (208, 356). The adenylylating and deadenylylating activities of this enzyme are sensitive to at least three allosteric effectors; GLN, $PIIUMP_3KG_3$, and $PIIKG_1$ (164, 178). The latter two species are the functionally most significant two out of the sixteen possible forms of the trimeric protein PII (280). This protein can be covalently modified by uridylyl-transferase (UTase) on each subunit (176). PII does not have any catalytic activity but solely functions as a regulatory protein (280). In order to be functional and to act as a substrate for UTase each of its subunits has to bind ATP (176). Given the range of cytosolic concentration of ATP and the dissociation constant of PII and ATP, PII is essentially always saturated with ATP (176). For convenience, the functional form of PII, *i.e.* $PIIATP_3$, will be denoted simply by PII throughout the remainder of the text. PII can in

addition bind α-ketoglutarate (176). The uridylylation and deuridylylation activity is allosterically regulated by glutamine and is independent of the α-ketoglutarate concentration (176).

PII not only affects the activity of ATase but also acts as the signal for the two-component signal transduction system NRI-NRII (177). The sensor of this system NRII (or NtrB) becomes activated by autophosphorylation of a histidine residue (174, 175, 282, 406). For the response regulator NRI (or NtrC) to become phosphorylated NRII~P has to transfer its inorganic phosphate onto an aspartate residue (295, 406). Also acetyl-phosphate can act as a phosphate donor (cf. 132). Dephosphorylation of NRI~P is carried out by the complex NRII-$PIIKG_1$ (177, 280). Spontaneous dephosphorylation of NRI~P occurs as well (132).

The phosphorylated form of NRI activates transcription of promoters that are part of the Ntr (Ntr: nitrogen regulated) regulon (281). This regulon comprises approximately a hundred genes (312). Their expression depends on the level of NRI~P (10). The *glnALG* operon is part of the Ntr regulon and is activated by NRI~P via one of its three promoters, *i.e.* *glnAp2*. It contains three structural genes *glnA* (GS), *glnL* (NRII) and *glnG* (NRI) (47). Two of the three promoters of the *glnALG* operon are upstream of glnA, *i.e.* *glnAp1* and *glnAp2*, and the third *glnLp* is upstream of *glnL* (281). Also upstream of *glnLp* a transcriptional terminator is located, which appears to be leaky (132). Transcription at *glnAp1* and *glnLp* occurs independently of NRI~P, is sigma factor σ^{70} dependent, and takes place when cells grown in nitrogen rich medium and is inhibited by NRI (132). At high levels of NRI and NRI~P, *i.e.* in ammonium limited medium, transcription starting at *glnAp1* and *glnL* is inhibited and is taking over by the more potent promoter *glnAp2* activated by NRI~P (281). Transcription from *glnAp2* makes use of an enhancer mechanism and of the sigma factor σ^{54} (or σ^N) (132). This mechanism involves the formation of a loop in the DNA that allows for two phosphorylated molecules of NRI, bound to the DNA at so-called enhancer sites at some distance from the *glnAp2* promoter, to make contact with the σ^{54}-RNA polymerase complex to the *glnAp2* promoter. Upon contact, the DNA is melted, open complex formation occurs, and transcription is started. It has been suggested that the σ^{54}-RNA polymerase complex is bound so tightly to the DNA that the *glnAp2* promotor can be considered occupied at all times in cells in which this complex is present (332). At higher levels of NRI~P, transcription of *glnAp2* is inhibited by NRI~P, possibly by its binding to two low affinity binding sites present between *glnAp2* and the enhancer binding sites (281). Hereby NRI~P inhibits the formation of the DNA loop and inhibits is own transcription and that of NRII and GS. This mechanism

presumably prevents accumulation of GS, NRI, and NRII. Transcription from the *glnAp1* operon of *glnA* has been shown to be activated by CRP-cAMP (CRP stands for cAMP receptor protein and is some times referred to as catabolite gene activator protein (CAP)), (132, 383). During growth on glucose (PTS-carbohydrates) the concentration of CRP-cAMP is low (383).

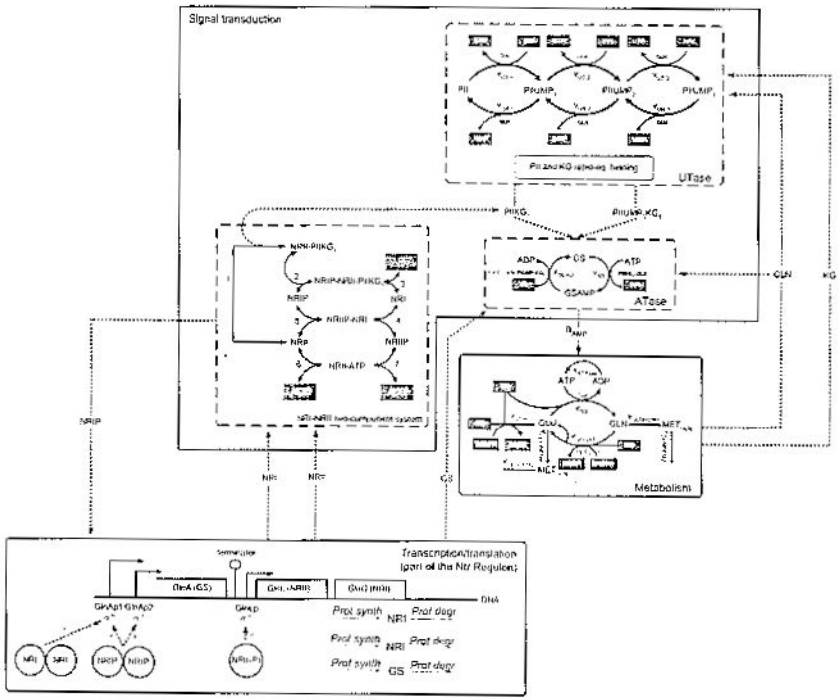

Figure 10.

Modular decomposition (on basis of the K matrix) of the regulatory network underlying ammonium assimilation in Escherichia coli. The network is decomposed into three modules, i.e. metabolism, transcription/translation, and signal transduction. The signal-transduction module consists of three interacting subnetworks, i.e. characterized by UTase, ATase, and NRI-NRII two-component signal transduction. The grey-boxed metabolites are assumed to be buffered by the remainder of metabolism. The intermediates displayed outside the modules are the communicating intermediates that govern the interconnections between the modules. The interactions between the modules are depicted as dashed arrows. Activators and inhibitors are depicted in bold and plain format, respectively. Abbreviations; GLN: glutamine, NH: ammonium, GLU: glutamate, n_{AMP}: the adenylylation state of GS, MET_{GLN} & MET_{GLU}: metabolites (e.g. amino acids) derived from glutamine and glutamate, respectively.

The transcription of GDH from *gdhA* is under the control of Nac (49, 124, 269). The expression of *nac* is activated by NRI~P (10). Transcription of *gltBD* encoding the two subunits of GOGAT is under the control of Lrp (96, 97, 311). GlnK, a homologue of PII, is expressed under conditions of severe ammonium starvation (12, 28, 382, 390). The roles of Nac, Lrp, and GlnK in ammonium assimilation will not be further discussed here (cf. (132, 311)).

The glutamine synthetase cascade that constitutes the metabolic regulatory machinery of ammonium assimilation has been intensively analysed since the seventies (132, 366). This network constitutes an ideal candidate for the construction of a detailed model, both from the perspective of its complexity and from the detail with which it has been experimentally assessed during the last decades. The kinetic properties of the major proteins have been studied and it appears that the majority of the regulatory interactions are known. Therefore, a silicon cell type of model of this network might approximate the physiological behavior of ammonium assimilation by *E. coli*. This model can assist in the further analysis of the network, which may lead to proposals for experiments and to the discovery of previously unanticipated regulatory phenomena.

4. Consequences of contemporary biology for the philosophy of biology

4.1 Scientific practice and philosophy

The biological sciences have evolved enormously in the previous century. The developments have not only had technological and scientific consequences. Many insights of modern biology also have had philosophical consequences. For instance, nowadays cells are considered in principle explicable in terms of physics, whereas in the late 19th and early 20th century many influential scientists argued that there exist a 'vital force' that makes living systems alive and which is not within reach of physical explanation (51). In addition, the development of nonlinear science and its influences on biology have had great consequences for the philosophy of biological systems (118, 278, 299, 384). It led to the awareness that, and it explained how, systemic properties that were not displayed by any of the components in isolation can arise out of nonlinear interactions among components of systems. In the next section, a simple example will illustrate this point. In particular, many of the properties that arise at the level of biological organization are not evident at lower levels of organization. For instance, function, evolution, adaptation, diauxic growth, and cooperativity can be considered properties that 'emerge' at the level of biological systems.

Nature is organized in a mereological (part-whole) manner: larger systems are

composed of smaller subsystems, which are composed of even smaller sub-subsystems. In addition, it is generally accepted that given the environment the systemic properties result from the properties of the parts and their interactions.

> 'The notion of a level implies a partial ordering, such that higher entities are composed of lower level entities, and, in a universe where reductionism is a good research strategy, the properties of higher level entities are predominantly best explained in terms of the properties and interrelations of lower level entities' (pg 680 in Wimsatt (422))

In principle, the decomposition of systems into parts can be repeated until we end up at the lowest level of elementary particles and their interactions. The consequences of this view have led to all kinds of questions that are of interest to both philosophers and scientists. Their answers may contribute to a better understanding of the organization and the functioning of complex systems and, in particular, the systems studied in biology.

The success of physics and the molecular biosciences have convinced many of us that living organisms are physical systems, *i.e.* that they can be understood in terms of theories about physical entities. Consider this quote from Richard Feynman:

> 'For example, life itself is supposedly understandable in principle from the movements of atoms, and those atoms are made out of neutrons, protons and electrons. I must immediately say that when we state that we understand it in principle, we only mean that we think that, if we could figure everything out, we would find that there is nothing new in physics which needs to be discovered in order to understand the phenomena of life. Another instance, the fact that the stars emit energy, solar energy or stellar energy, is presumably also understood in terms of nuclear reactions among these particles. All kinds of details of the way atoms behave are accurately described with this kind of model, at least as far as we know at present. In fact, I can say that in the range of phenomena today, so far as I know there are no phenomena that we are sure cannot be explained this way, or even that there is deep mystery about.' (Richard Feynman, pg 145 in (107))

Statements like this one have given rise to many discussions about the status of higher-level properties; *e.g.* about whether they can be considered emergent, about whether the sciences that study them can preserve their autonomy from physics, and about whether they are reducible to physics (– and about what reductions mean in this context?).

4.2 Nonlinear science and complexity

In 1948, Weaver noticed that, so far, science had been mostly successful in the construction of theories about systems either with a small number of variables that can be treated with mechanics (*e.g.* the motion of two billiard balls) or with an enormous number of variables that can be treated with statistical mechanics (*e.g.* average properties of ideal gases) (405). Weaver terms the kind of complexity displayed by the latter kind of systems: "disorganized complexity". The middle region of phenomena, which involve systems with only a moderate number of variables that are interrelated, result from "organized complexity". It was organized complexity that has been left mostly untouched in the natural sciences before the fifties of the last century (405). 'They are all problems which involve dealing simultaneously with a *sizeable number of factors which are interrelated into an organic whole.* They are all, in the language here proposed, problems of *organized complexity.*' (pg 539 in Weaver (405)). (In present day terminology, what Weaver meant with "interrelated" and with "organic whole" comes close to "nonlinearly related" and "complex system" in more modern terms, respectively.) Since 1948, science has made a great leap forward with the introduction of computers, with the development of theories on the dynamics of nonlinear systems, and with the mechanistic elucidation of a number of complicated biological systems. Similar analysis of the profound differences between 'old' physics and nonlinear systems can be found in the works by Prigogine & Stengers (299), Anderson (6), Monod (265), and Krebs (228).

4.3 Appearance of new behavior

Nonlinearity and its consequences become apparent in systems that are displaced from thermodynamic equilibrium (118, 278). To illustrate the concept and the consequences of nonlinearity a simple example taken from Nicolis & Prigogine (pg 59-60 in Nicolis & Prigogine (278)) will be analyzed. It consists of the two chemical reactions given below.

$$A + 2X \underset{}{\overset{1}{\longleftrightarrow}} 3X$$

$$X \underset{}{\overset{2}{\longleftrightarrow}} B$$

The former reaction involves the autocatalytic synthesis of three molecules of X out of one molecule of A and two molecules of X. The latter reaction involves the conversion of X into B. Both reactions are considered reversible and the concentration of A and of B shall be assumed fixed. When defined in this manner, the system contains only the concentration of X as the dynamic variable. The change in X at time t is given by its differential equation,

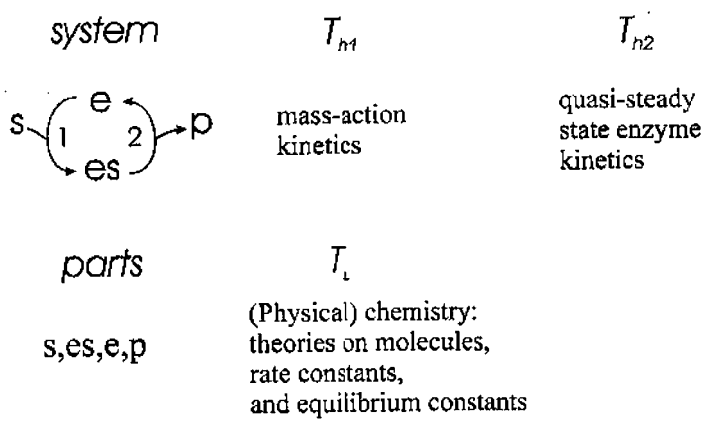

Figure 12.

Two levels are depicted. A higher level that represents a system composed of interacting parts in some environment at given physicochemical conditions. A lower level that describing the parts in isolation under the same physicochemical conditions. A lower-level theory T_L (physical chemistry) describes the properties of the parts. Two higher-level theories T_{H1} and T_{H2} describe the properties of the system. T_{H2} can be obtained from T_{H1} by assuming a quasi-steady state for the concentration of the enzyme-substrate complexes (cf. 62, 353)(see also Figure 2).

4.4 Reductionism

The concept of reductionism will be explained on the basis of an example taken from biochemistry, *i.e.* the kinetics of enzyme catalysis (Figure 12). This figure displays a lower level that contains the parts and a theory T_L (physical chemistry) that describes the properties of the parts only and a higher level that is composed of the same parts engaging in interactions. The net effect of the interactions in the system at the higher level is that the product *p* is formed from the substrate *s* (and vice versa) with enzyme *e* acting as the catalyst. The system property of interest is the rate of change of *p*, *i.e.* d*p*/dt. With respect to this system property, the behavior of the system can be described by two theories, T_{H1} and T_{H2}. Theory T_{H1} is a mass-action description of the reactions taking place in the system. Theory T_{H2} is a description of the enzyme using quasi-steady state enzyme kinetics of the reactions taking place in the system. When do philosophers speak of reduction in this context? Is there a difference between reduction of theories, *i.e.* the reduction T_{H2} to T_{H1} or the reduction T_{H1} to T_L, and the reduction of phenomena (such as d*p*/d*t*)?

One of the classical works on reduction comes from Ernest Nagel, who focused on reduction in his analysis of scientific progress (271). Nagel was interested in theory

reduction, because he considered it to be a primary event within the progress of all sciences. Nagel formulated formal rules for the reduction of theories based on Hempel's deductive-nomological model of explanation. Nagel distinguished between homogeneous and nonhomogeneous reduction (270). The latter meant that the theory to be reduced needed amendments in order to allow for the reduction, while in the former case the reduction was direct. Such amendments have been termed 'bridge laws'. Bridge laws are necessary in nonhomogeneous reductions. Most reductions are nonhomogeneous. A classical example of reduction of theories is the reduction of macroscopic thermodynamics to statistical thermodynamics, *e.g.* the identification of the temperature of a system with the average kinetic energy of its molecules.

Many are opposed to Nagel's intertheoretic model for reduction and for many different reasons. For our purposes, the most interesting critique is that the model of intertheoretic reduction of Nagel is rarely found in scientific practice (*cf.* (422)). Wimsatt (422) (and many others) has tried to give an account of reduction that does justice to scientific practice.

Wimsatt distinguishes between two forms of reduction: (i) successional or intralevel reduction and (ii) explanatory or interlevel reduction (422). Intralevel reduction has to with the relationship between T_{H1} and T_{H2}. These two theories can describe the same systemic property (in this case d*p*/d*t*). They can therefore be considered 'competing' theories. Wimsatt argues that theory succession or theory elimination – frequently associated with reductionism – is only meaningful in the case of intralevel reductions. In our example, T_{H2} can be derived from T_{H1} if it is assumed in T_{H1} that the concentration of *es* remains effectively constant on the time scale of the conversion of *s* into *p*; this is known as the quasi-steady state assumption (62).[4] However, many biochemists would not consider this an elimination of T_{H2}: descriptions that make use of quasi-steady state enzyme kinetics are used at least as frequently as descriptions based on mass-action kinetics (T_{H1}).

Interlevel or explanatory reduction is the form of reduction more scientists speak of. Explanatory reduction is generally not in terms of reduction of theories but it involves the search for a mechanism that can account for a systemic behavior in terms of the interacting parts of the system (422). This form of reduction is in contemporary philosophy referred to as mechanistic explanation (120, 240).

[4] It seems that this reduction obeys the formal model of Nagel, but it will not be considered here any further. This example of reduction would qualify as a nonhomogenous reduction; the quasi-steady state assumption of *es* then serves as the bridge law.

4.5 Irreducibility and emergence

One way to question the reductionist tendency towards reduction of phenomena in terms of more fundamental theories, such as theories of physics, would be to show that there exist higher-level phenomena – higher than the level of physical phenomena – that cannot be reduced to physics. This would mean that such higher-level phenomena necessitate laws that are irreducible to the laws of physics. The putative existence of such irreducible phenomena has been fiercely debated in the philosophy of science community and this discussion resurfaces now and then. One interpretation of emergent phenomena is that they are phenomena which are intrinsic to particular levels of organization and that they cannot be reduced on the basis of the phenomena that take place on lower levels of organization alone (*e.g.* (205)). However, perhaps also weaker forms of emergence can be defined which are still compatible with reductionism without trivializing emergence, and still refer to a class of systemic properties in a meaningful way.

Stephan (370, 373) discussed various definitions of emergence. His 'weak emergentism' is compatible with property reduction. A property that is weakly emergent has to comply with three conditions: i. *physical monism*: 'Entities existing or coming into being in the universe solely consist of material parts. Likewise, properties, dispositions, behaviors, or structures classified as emergent are instantiated by systems consisting exclusively of physical parts.' (pg 50 in Stephan (373)), ii. *systemic properties*[5]: 'emergent properties are systemic properties. A property is a systemic property if and only if a system possesses it, but no part of the system possesses it.' (pg 50 in Stephan (373)), and iii. *synchronic determination*: 'A system's properties and dispositions to behave depend nomologically on its micro-structure, that is to say, on its parts and their arrangement'. There can be no difference in the systemic properties without there being some differences in the properties of the system's parts or their arrangement.' (pg 51 in Stephan (373)). All stronger forms of emergence defined in Stephan (373) comply to the conditions for weak emergence and to some additional conditions. A weakly emergent property becomes synchronically emergent if it is irreducible: 'A systemic property [organizational property] is irreducible if (a) it is neither micro- nor macroscopically behaviorally analyzable, or if (b) the specific behavior of the system's components, over which the systemic property supervenes, does not follow from the component's behavior in isolation or in other (simpler) constellations.' (pg 52-53 in Stephan (373)). In contrast to synchronically emergent properties, diachronically emergent properties have a temporal component, *e.g.* a weakly emergent property becomes diachronically emergent if it could not have predicted before its first observation of instantiation (373). Diachronically emergent properties will not be further discussed here.

[5] So far systemic properties have been defined rather loosely. To prevent any confusion, the condition of systemic properties will be referred as the condition of *organizational properties* from now on. Organizational properties were termed "reducible characteristics of the order" by Broad (pg 78 in (41)). Broad termed properties had by the systems and had by its parts "ordinally neutral properties" (pg 78 in (41)).

At present status many concrete examples of biochemical reaction networks exist that have been characterized molecularly to such an extent that the construction of detailed kinetic model becomes possible. In addition, the systemic behavior of those models can be compared to *in vivo* behavior to investigate whether our molecular knowledge can account for the systemic behavior. Such kinetic models are ideal vehicles to investigate the present status of reductionism and emergence in the analysis of cell biology.

5. Outline of this thesis

The central problem addressed in this thesis is how properties and behaviors of living cells can be understood in terms of the properties and behaviors of their constituent macromolecules. The thesis consists of two parts. The first part is concerned with the philosophical aspects and the second part with the systems-biological aspects of the central problem.

The first part deals with philosophical issues in relation to research of living systems in terms of the properties of and interactions between their components. This part consists of Chapter 2 & 3 of the thesis. **Chapter 2** is concerned with the philosophical issue of mechanistic explanation. It starts with a review of the major developments in the philosophy of science that have led to the present theory of mechanistic explanation. It analyzes to what extent the current theory of mechanistic explanation can accommodate the type of mechanistic explanations found in systems biology. It offers suggestions to further develop the theory of mechanistic explanation. **Chapter 3** offers an account of emergence and a new definition of emergent properties applicable to the analysis of biochemical reaction networks. It starts with a discussion of definitions of emergent properties offered by other philosophers. The definition of emergent properties proposed in this chapter is compatible with a mechanistic explanation of emergent properties. In this chapter it is also explained why mechanistic explanations of the behavior of living cells in terms of their molecules is done in terms of the behavior of the molecules within the system rather than in terms of the properties of the molecules that can be determined in isolation.

The second part is concerned with the systems-biological analysis of biochemical reactions networks. This part consists of Chapters 4 to 6 of this thesis. **Chapter 4** presents a detailed kinetic model of the metabolic regulation of ammonium assimilation in *Escherichia coli*. In this chapter, the model is analyzed for its steady and transient state characteristics as a function of the nitrogen and the carbon status of the cell. A method is developed to dissect the transient regulation of the rate of glutamine synthetase into the contributions of the regulators present in the system. The effects of gene expression of glutamine synthetase, glutamate dehydrogenase, and glutamate synthase on the steady-state ammonium assimilation flux are discussed. **Chapter 5** presents modular response analysis, an extension to modular control analysis, as a method to analyze systemic responses of modular biochemical networks in terms of the strength of the interactions among their constituent modules. The modules considered communicate by way of regulatory influences only, and not by way of mass flow. The method allows for a significant reduction of complexity by offering the possibility to focus only on the intermediates that engage in intermodular interactions. The remaining intermediates, which are only involved in intramodular interactions, can be omitted from the analysis. Because of its modular character, modular response analysis also allows for a gradual increase in the complexity of the analysis of the modular system by offering a method to sequentially 'open' more modules as the analysis progresses. The method is illustrated by calculation of the modular interaction strengths in the kinetic model presented in Chapter 4. **Chapter 6** offers a definition of robustness of biochemical reaction network properties in terms of metabolic control analysis. It provides three methods that explain the occurrences of robustness of system properties of biochemical reactions networks. In addition, interesting cases of robustness from a cell biological perspective are distinguished from trivial cases of robustness on the basis of the presence or absence of a molecular mechanism that actively brings about robustness. The robustness of the ammonium-assimilation flux with respect to changes in the ammonium concentration observed in Chapter 4 is explained in terms of the modular interaction structure of ammonium assimilation in *E. coli*.

Chapter 7 offers a general discussion of the thesis. It discusses the main results in relation to the central problem that is addressed in this thesis.

2

Mechanistic explanation

'The goal of understanding the world is a theoretical goal, and if the world is a machine – a vast arrangement of nomic connections – then our theory ought to give us some insight into the structure and workings of the mechanism, above and beyond the capability of predicting and controlling its outcomes ... Knowing enough to subsume an event under the right kind of laws is not, therefore, tantamount to knowing the how and why of it. As the explanatory inadequacies of successful practical disciplines remind us: explanation must be more than potentially-predictive inferences or law-invoking recipes.'
(Railton, P., A Deductive-Nomological Model of Probabilistic Explanation, Philosophy of Science, 45, 206-226, 1978)