CHAPTER 11

SYSTEMS BIOLOGY AND THE RECONSTRUCTION OF THE CELL: FROM MOLECULAR COMPONENTS TO INTEGRAL FUNCTION

FRANK J. BRUGGEMAN[1], SERGIO ROSSELL,
KAREN VAN EUNEN, JILDAU BOUWMAN,
HANS V. WESTERHOFF[1] and BARBARA BAKKER
Free University, The Netherlands
[1] *University of Manchester, UK*

Table of Contents

E. Bertrand and M. Faupel (eds.), Subcellular Proteomics, 239–262.

1. SYSTEMS BIOLOGY

1.1. Introduction

Molecular biology was one of the most successful scientific disciplines of the previous century. It matured into an enormous undertaking involving many tens of thousands of scientists all over the world and has spawned a huge industry involved in medicine and (bio-) technology. The discoveries of molecular biology have been driven to a large extent by the development of new technologies enabling the scientists to identify and investigate new molecular phenomena in cells.

Molecular biology began to take shape in the 1940s. In those days the "aim" of the field became transparent: molecular biology was to discover and characterize the molecular constituents of cells (Stent 1968; Hess 1970). And that is certainly what it has been doing very successfully over the last 60 years or so.

We are now at the point in time where our understanding of the properties of the molecular constituents of cells is of sufficient detail to start studying the properties of networks composed of interacting molecules. This requires an approach different form that of molecular biology, involving the application of different techniques and methods. This new field in biology has been emerging in the last few years and is referred to as "Systems Biology" (Ideker et al. 2001a, b; Alberghina and Westerhoff 2005).

Systems biology studies how the functional behaviour of biological networks arises from the interactions between the molecules of which they consist. Accordingly, the introduction of systems biology involved a change in focus: from molecule to network. This has become possible at a large scale only recently, thanks to the enormous knowledge that molecular biology, biochemistry and cell biology have acquired about the molecular organization of cells, and thanks to genome sequencing and a set of techniques enabling functional genomics. There are two forms of systems biology: bottom-up (investigates how the molecular interactions between a set of known components gives rise to functional behaviour) and top-down (uses system level data to identify and characterize underlying patterns of interactions). As we shall see the type of research associated with the former form of systems biology is new (thanks to a number of developments in mathematical biochemistry), as is the applications of top-down systems biology. Systems biology may rightfully be called a new discipline in science by virtue of the fact that it differs essentially from both physical chemistry and biology.

The aim of systems biology is to come to understand general principles underlying the emergence of the behaviour of cells from their molecular organization – it addresses how molecules jointly bring about cellular behaviour. This level of understanding should enable more rational approaches to current challenges encountered in medicine and biotechnology such as drug design and product formation. When systems biology succeeds we should have complete understanding of the molecular functioning of cells. Although systems biology may seem to represent a natural continuation of molecular biology, it requires a transition in biological thinking to enter this new phase towards understanding cells and organisms.

1.2. Two Approaches to Systems Biology

As was mentioned above, systems biology studies how cellular phenomena derive from the interactions between the molecular constituents of cells. This means that studies have to be done on the level of the entire cell, for example, apoptosis, cell division, and adaptation, as well as on the molecular level, for example, studies of protein–protein interactions, cellular location, and molecular transport. Most importantly, and this is perhaps the most challenging aspect of systems biology, the corresponding two datasets have to be integrated to the extent that it allows understanding of how the joint molecular behaviours – the latter dataset – underly the cellular behaviour – the former dataset.

Two strategies may be followed. *Top-down systems biology* starts with the system-level data to identify and characterize underlying molecular mechanisms. *Bottom-up systems biology* starts from the other end, at the level of the molecular mechanisms, and investigates how the latter jointly bring about cellular phenomena. Top-down systems biology is used more often for "discovery" studies, for example, to predict the topology of protein–protein interaction networks, or (presumed) transcription factor networks (target genes for transcription factors) from data concerning all molecules of the corresponding subtype of the entire cell (or a large proportion thereof) (Ideker et al. 2001a, b). This involves more than, for instance, the use of microarrays or yeast two-hybrid screens. It also requires the use of new computational methods, often statistical, to obtain from those datasets network statistically significant topologies that will function as predictions. Top-down systems biology tends to be more associated with studies of poorly-characterized networks of cells. Bottom-up systems biology starts with all the knowledge about one particular subnetwork in a cell and considers the properties of its molecules as well as the processes that underlie their interactions. It then investigates: (i) how its behaviour derives from its internal molecular organization, (ii) how it influences the remainder of the cell and *vice versa* and (iii) what the function(s) are of this subnetwork inside the cell.

Both approaches to systems biology rely on mathematical methods and modelling, as well as on quantitative and precise experimentation. However, the types of mathematical and experimental approaches used differ between the two. Bottom-up systems biology uses either detailed kinetic models describing the processes taking place between molecules in terms of experimentally determined kinetic properties (if available) when it wishes to identify actual mechanisms of emergence of functionality, or "core" models when it examines what new properties might arise from certain interactions (Bakker et al. 1997; Kholodenko et al. 1999; Rohwer et al. 2000; Teusink et al. 2000; Hoefnagel et al. 2002; Hornberg et al. 2005). It relies on knowledge about the topology of the molecular network and the properties of molecules. Top-down systems biology is applied to actually determine topology and occasionally kinetic information possibly in a data driven, hypothesis free fashion. Below we will consider kinetic models, and how they are used in bottom-up systems biology, more elaborately.

1.3. Two Historical Origins of Systems Biology

Systems biology applies sophisticated mathematical and experimental methods to analyse the functioning of molecular networks in cells (Brenner 1999; Kitano 2002; Hood 2003; Bruggeman et al. 2005a; Kriete and Eils 2005). Accordingly, not only developments in molecular biology contributed to the growth of systems biology but also developments in the analysis of nonlinear systems in non-equilibrium physics and mathematics. Whereas developments in the analysis of three-dimensional structures of macromolecules, using crystallography or NMR, were of minor importance for the concepts of system biology, developments such as those of enzyme kinetics, the chemiosmotic theory, allosteric enzyme regulation, covalent modification, regulation of the *lac* operon, were of prime importance. Bioenergetics, enzymology, and biophysics have therefore been influential disciplines for (bottom-up) systems biology. The experimental and theoretical analysis of the stationary and dynamic states of nonlinear systems operating under non-equilibrium conditions is also an important root of systems biology. This includes non-equilibrium thermodynamics, network thermodynamics, the analysis of periodic and stochastic phenomena, control theory, modelling, bifurcation theory, and nonlinear time series analysis. Even though the latter methods have been applied to biology prior to the emergence of systems biology, with exceptions like mosaic non-equilibrium thermodynamics and metabolic and hierarchical control analysis, they have rarely been applied in a way that referred to the macromolecular components of the system, using both experimental data on molecular properties and on the system level. More about the historical roots of systems biology can be found in Westerhoff and Palsson (2004).

2. OVERVIEW OF APPROACHES TO UNDERSTAND CELL FUNCTION BASED ON MOLECULAR INTERACTIONS

2.1. Topological Analysis

Depending on the purpose – "which biological question is to be answered?" – different kinds of modelling can be distinguished (Bruggeman 2005). Very often, the choice of the modelling approach depends also on the quality and extent of molecular information available about the interactions out of which the network is composed. The description that is most qualitative, but nevertheless can answer important biological questions, considers the network as a graph. In this approach, all the molecules in the network represent nodes and their interactions represent edges. This is the only information that is used in this modelling approach. Often such graphs are described in terms of an adjacency matrix denoted by **A**. The rows and columns denote the identity of the interacting molecules and an inhibiting interaction from molecule j to k is denoted by -1 in the j-th row and k-th column of **A**, that is, $a_{jk} = -1$. A "1" would denote an activatory interaction. In some cases, the direction of the interaction is not taken into account, making **A** symmetric, that is, for j and k we have $a_{jk} = a_{kj}$. Research questions concerning the strengths of the interactions or

reaction rates (fluxes) cannot be answered with this modelling approach. In the past few years, methods based on graph theory have been used to study the robustness and the modular, scale-free and small-world organization of genetic, protein–protein interaction, signalling and metabolic networks (Watts and Strogatz 1998; Fell and Wagner 2000, 2001; Milo et al. 2002; Ravasz et al. 2002; Spirin and Mirny 2003; Pereira-Leal et al. 2004). Many analytical and numerical tools have been developed to study such networks (Albert and Barabasi 2002; Itzkovitz et al. 2003; Newman 2003).

In this approach only static topological properties of networks can be investigated; such as the number and length of the paths between two molecules, the number and length of cycles, the enumeration of occurrences of particular subnetwork structures, and how those properties scale with the size of the network. This type of modelling we shall refer to as *topological network modelling*.

2.2. Stoichiometric Analysis

The next type of modelling we shall discuss is *stoichiometric modelling* (Schilling et al. 1999; Price et al. 2004). In this approach, it is emphasized that all molecular interactions are in fact reactions having particular stoichiometric coefficients. The latter are the numbers (often integers) preceding the reactants in a reaction, for example, the numbers in: $2A + B + 3C \leftrightarrow 2AC + BC$. This approach appears much more relevant for the analysis of metabolic networks than for genetic and signalling networks where such stoichiometries are absent or ill-defined (for promising exceptions see (Allen and Palsson 2003; Papin and Palsson 2004a, b; Papin et al. 2005)). Again this analysis may be performed in terms of matrices. The stoichiometric matrix **N** has as many rows as there are independent molecule concentrations (or rather variable intermediates engaging in reactions) in the network and as many columns as there are reactions in the network. Then the stoichiometric coefficient of the j-th intermediate in the network, say glutamate, for the k-th reaction is the system, say glutamate synthase, is 2, whereas glutamine, the l-th intermediate, has a -1 in the kth-column of **N**, that is, $n_{jk} = 2$ and $n_{lk} = -1$.

One type of stoichiometric analysis concerns the discovery of moiety-conservation relationships in the network (Reder 1988; Famili and Palsson 2003; Nikolaev et al. 2005; Imielinski et al. 2006). These linear relationships represent sums of concentrations of intermediates in the network that remain constant on the time scale of interest, because there are only reactions that interconvert these particular intermediates. For instance, the following moiety-conservation relationship may pertain: $[P_i] + [AMP] + 2[ADP] + 3[ATP] +$ [other phosphate containing molecules] $= [P]_{total}$ constant.

A more popular form of stoichiometric analysis is the analysis of flux distributions that are consistent with system steady state (Note that in the terminology of metabolic modelling, the rate of a reaction at system steady state is referred to as a flux.) This type of analysis can be done directly on the **N** matrix because of its central role in the description of the mass balances of all the variable intermediates in a network.

A mass balance is an equation for each variable intermediate in the network, say x, that describes at a particular moment in time the rate of change in this variable, that is, $\mathrm{d}x/\mathrm{d}t$, in terms of the difference between the rates of all the reactions producing x and all the reactions consuming it. If we denote rates by v then we obtain for each of the intermediates, entries in the state vector $x = \{x_1, \ldots, x_j, \ldots, x_m\}$, the following mass balance:

$$\frac{\mathrm{d}}{\mathrm{d}t}x_j = \sum_k n_{jk} \cdot v_k \tag{1}$$

In matrix format we obtain (denoting a vector in bold small font and a matrix in bold capital font):

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{N}\mathbf{v} \tag{2}$$

In steady state, we denote the rate vector $\mathbf{v}$ by the flux vector $\mathbf{J}$ (this will be the only vector denoted by a capital letter) to obtain:

$$\mathbf{0} = \mathbf{N}\mathbf{J} \tag{3}$$

Many different flux vectors will satisfy this equation. All of those solutions are linear combinations of the independent fluxes in the network. The coefficients to appear in those linear combinations derive from the nullspace or kernel matrix of $\mathbf{N}$ denoted by $\mathbf{K}$ (Schuster et al. 1999). Reordering $\mathbf{J}$ with the independent fluxes $\mathbf{J}^{\mathrm{I}}$ on top of the dependent fluxes $\mathbf{J}^{\mathrm{D}}$ one obtains:

$$\mathbf{J} = \mathbf{K}\mathbf{J}^{\mathbf{I}} \tag{4}$$

Each column of $\mathbf{K}$ represents a flux mode, a set of fluxes through the reaction steps in the network that are consistent with a system steady state. The choice of $\mathbf{K}$ is not unique. To solve this and to make the definition of flux mode biologically more interesting and applicable to bioengineering and pathway analysis, elementary modes and extreme pathways have been developed, both corresponding to choices of $\mathbf{K}$ (Schilling et al. 2000; Schuster et al. 2000; Papin et al. 2004) that are extreme in terms of simplicity. These choices have been useful in analysing the capabilities of the genome-scale metabolic network, to address characteristics such as possible yields, robustness, the viability of mutants and the redundancy of pathways (Price et al. 2004).

Another application of the analysis of the stoichiometric matrix is flux balance analysis (Edwards et al. 2002). Often the number of fluxes in the system exceeds the number variable metabolites making equation (3) an underdetermined set of linear equations, that is, many different combinations of fluxes are consistent with system steady state. One approach is to measure the fluxes that enter and exit the cell. Because intracellularly there are many redundant pathways, this does not enable one to determine all fluxes. Isotope labelling may help then (Wiechert 2002). Another approach to then find a smaller number of solutions is to postulate that the solution should satisfy an additional objective. This objective is taken to be associated with optimal functioning of the network, for instance maximization of some flux or combination

of fluxes (e.g. growth rate or flux ratio). This type of approach is termed flux balance analysis. When investigating genome-scale networks the number of solutions may then still run into the hundreds but at least such numbers are manageable (Mahadevan and Schilling 2003; Reed and Palsson 2004).

2.3. Kinetic Modelling

None of the methods so far were able to deal with dynamics of intracellular networks. They were not able to describe the changes in the concentrations of the network intermediates as function of time upon perturbations made to the network, such as the addition of nutrients, growth factors, or drugs. This is what kinetic modelling does. A kinetic model starts from equation (2) by substituting rate equations into the rate vector. Rate equations describe the dependence of a rate of a reaction in the network with respects to its substrates, products, and effectors by the identification of the enzyme mechanism and the parameterisation of its kinetic constants. An example of a rate equation is the following two substrate (s_1 and s_2) and two product (p_1 and p_2) reaction with the non-competitive inhibitory effect of x:

$$v = \frac{V^{+}_{\mathrm{MAX}} \cdot \frac{s_1 \cdot s_2}{K_{s_1} \cdot K_{s_2}} - V^{-}_{\mathrm{MAX}} \cdot \frac{p_1 \cdot p_2}{K_{p_1} \cdot K_{p_2}}}{\left(1 + \frac{x}{K_x}\right)\left(1 + \frac{s_1}{K_{s_1}} + \frac{s_2}{K_{s_2}}\right)\left(1 + \frac{p_1}{K_{p_1}} + \frac{p_2}{K_{p_2}}\right)} \tag{5}$$

V^{+}_{MAX} and V^{-}_{MAX} denote the maximal rate of the reaction in the forward and the backward directions respectively. The "K" constants reflect the effective affinity of the enzyme for its substrates, products and its one effector. All those constants can be determined experimentally. Most often those experiments are carried in cell free extract or on isolated enzymes. More on enzyme kinetics, including more complicated cases involving multimeric enzymes, can be found in Segel (1993) and Cleland (Cleland 1963a, b, c). This is a discipline in itself, even though much smaller nowadays than in the 1960s and 1970s. Supplemented with initial conditions, the kinetic model, with the parameterized rate equations incorporated, allows the calculation of the temporal behaviour of network.

Not all kinetic models derive their kinetic parameters and the exact enzyme mechanism from experiment. In those cases, the network may still be precise but the descriptions of the enzymes are then phenomenological. On the other end of the scale, "Silicon Cell" models are defined as to contain the experimentally confirmed network structure, enzyme mechanisms, and kinetic parameters (see www.siliconcell.net). Models that serve illustrative purposes or more qualitative purposes are core models and those have simplified network structures, simplified enzyme mechanisms, and guessed or fitted kinetic parameters. It all depends on the purpose of the model whether a model can be judged as good or bad. Silicon Cell models are especially promising because they contain the current status of understanding and should therefore be more easily integrated with experiment. They may assist in pinpointing gaps in our understanding, offering hypotheses, performing *in silico* "experiments," predicting

drug targets, and integrating various sources of biological data (fluxes, kinetic constants and concentration of mRNAs, proteins, and metabolites). Silicon cell models come as close as possible to offering a complete description of cells. The Silicon Cell programme will be discussed in more detail in Section 3.

2.4. Metabolic Control Analysis

A powerful tool to understand the behaviour of cells, pathways or models thereof is Metabolic Control Analysis. It was pioneered by Heinrich and Rapoport (1974) and Kacser & Burns (1973) in the seventies for steady-state metabolic networks with early experimental application by Flint et al. (1980, 1981) and Groen et al. (1982). MCA relates changes in systemic properties as quantified by response or control coefficients to properties of enzymes in terms of elasticity coefficients. MCA has later been extended to address branched and cyclic pathways with and without moiety conservation (Fell and Sauro 1985; Hofmeyr et al. 1986; Sauro et al. 1987; Reder 1988; Fell and Sauro 1990; Kacser et al. 1990; Sauro and Kacser 1990; Kholodenko et al. 1994b, 1995b; Sauro 1994, 1995 Westerhoff and Kell 1996), concentration control (Westerhoff and Chen 1984), energy coupling (Westerhoff et al. 1987), control of generalized variables (Schuster 1996), control of transition times (Easterby 1990; Melendezhevia et al. 1990), systems involving quasi-equilibrium reactions and time-scale separation (Delgado and Liao 1995; Kholodenko et al. 1998); oscillatory systems (Kholodenko et al. 1996, 1997a; Ingalls 2004; Sauro and Ingalls 2004), signaling networks (Kholodenko et al. 1997; Bruggeman et al. 2002), channeling (Kholodenko et al. 1994), intra-enzymatic processes (Kholodenko and Westerhoff 1994), hierarchical networks with gene expression, signaling, and metabolism (Westerhoff et al. 1989; Kahn and Westerhoff 1991; Jensen et al. 1999; Hofmeyr and Westerhoff 2001; Bruggeman et al. 2002; Snoep et al. 2002), modular biochemical networks (Schuster et al. 1993), reaction-diffusion systems (Brown and Kholodenko 1999; Peletier et al. 2003), and transient trajectories (Acerenza et al. 1989; Heinrich and Render 1991; Ingalls and Sauro 2003). It has been applied frequently to the experimental analysis of biochemical networks (Groen et al. 1982; Ainscow and Brand 1999a, b). In this section we will outline the theory only briefly. For more comprehensive reviews, see (Westerhoff and Van Dam 1987; Heinrich and Schuster 1996; Fell 1997).

The effect of a change of the activity of an enzyme i on the steady-state flux J is quantified by the corresponding flux control coefficient C_i^J. C_i^J is the percentage change of the flux J upon a 1% change of the enzyme activity v_i causing the change of the flux. Or, in more precise mathematical terms:

$$C_i^J \frac{(\mathrm{d} \ln J / \mathrm{d}p)_{ss}}{(\partial \ln \mathrm{v_i} / \partial \mathrm{p})_{\mathrm{reaction}_i}}$$

in which p is a parameter that only acts on enzyme i, such as an inhibitor concentration or an enzyme concentration. The subscript *ss* and differential "d" indicate that the

change of the flux is analysed after relaxation of all system variables to a new steady-state. The subscript *reaction i* and derivative "∂" indicate that only the change in the local rate of enzyme i is considered, immediately after the change of p at constant values of all system variables, such as the concentrations of substrates, products and effectors of enzyme i.

In simple metabolic pathways, with only one independent flux, the value of C_i^J is usually between zero and 1. Partial inhibition of an enzyme with zero flux control does not affect the flux, while an enzyme with a flux control coefficient of 1 is completely rate-limiting. Flux control coefficients between 0 and 1 have often been found (Groen et al. 1982; Ruijter et al. 1991; Snoep et al. 1996). In ideal metabolic pathways, for example, in the absence of metabolite channelling and coenzyme sequestration (Kholodenko et al. 1995), the sum of the flux control coefficients of all enzymes in the pathway must equal 1:

$$\sum_i C_i^J = 1$$

Control coefficients are properties of a metabolic pathway as a whole. The kinetic properties of individual enzymes that are relevant for control are expressed in terms of their elasticity coefficients. An elasticity coefficient $\varepsilon_{X_j}^i$ quantifies how strongly a metabolite concentration X_j affects an enzyme rate v_i directly, at constant concentrations of all other metabolite concentrations:

$$\varepsilon_{X_j}^i = \frac{\partial \ln v_i}{\partial \ln X_j}$$

Connectivity theorems allow to relate the control coefficients (systemic properties) to the elasticity coefficients (properties of the network's enzymes individually as if in isolation) (Westerhoff and Van Dam 1987; Heinrich and Schuster 1996; Fell 1997). The connectivity theorems have given us a strong insight into the functioning of metabolic pathways. For example, it follows directly from these theorems that enzymes that are very sensitive to the concentrations of metabolites, such as substrates, products and allosteric effectors, tend to have little control over the flux. This is illustrated by overproduction of phosphofructokinase in bakers' yeast, an enzyme often referred to textbooks as "rate-limiting." Yet, overproduction of phosphofructokinase does not lead to a significant flux increase, since the cell compensates by lowering the level of its allosteric effector fructose 2,6-bisphosphate (Schaaff et al. 1989; Davies and Brindle 1992).

2.5. Control versus Regulation

In the biological and biochemical literature the terms *control* and *regulation* are often used loosely as synonyms. In the precise analysis that is applied in systems biology, however, there is a need for a stricter definition of these terms.

We use the term *control* only in the sense in which it is used in Metabolic Control Analysis (Section 2.3). An enzyme is said to control a metabolic flux or a metabolite

concentration if activation of this enzyme causes a pronounced effect on this flux or concentration. In this sense questions about control are what-if questions. *If* the cell were to increase the concentration of a controlling enzyme or *if* a scientist were to add an inhibitor of this enzyme, the flux or concentration should change. Control coefficients do not tell us whether the cell uses this potential when it responds to environmental changes.

Regulation indicates which mechanisms the cell actually uses to effectuate changes or robustness of cellular functions. Different aspects of regulation are the (i) tendency of living organisms to respond to their environment, (ii) the internal communication between different parts of cells or organisms and (iii) the maintenance of a relatively constant internal state upon external perturbations (Kahn and Westerhoff 1993). The regulatory strength (Hofmeyr et al. 1993; Kahn and Westerhoff 1993; Hofmeyr 1995) is defined as:

$$^{i}R_A^J = C_i^J \cdot \varepsilon_A^1$$

It quantifies the effect of a perturbation of an external effector A (e.g. the concentration of a nutrient) on the flux J via enzyme i. If A effects different enzymes, perturbations of A can effect J via different routes and the regulatory strengths quantify the relative importance of each of these regulatory routes. A different, but related approach was taken by Westerhoff and Ter Kuile (ter Kuile and Westerhoff 2001). They developed Regulation Analysis, a theory that quantifies the relative importance of flux regulation via metabolism and via gene expression (Section 2.5).

In summary, *control* tells us about the buttons of which the cell disposes and *regulation* tells us which of these buttons the cell actually presses (through the elasticities of the components).

2.6. Regulation Analysis

Because enzymes are catalysts (and not substrates), enzyme rate equations are usually of the shape:

$$v = v(e, \mathbf{X}, \mathbf{K}) = f(e) \cdot g(\mathbf{X}, \mathbf{K})$$

in which v is the rate, e is the enzyme concentration, $\mathbf{X}$ is a vector of concentrations of substrates, products and other metabolic effectors, and $\mathbf{K}$ is a vector of constants parameterizing the strength with which the enzymes interact with their substrates, products and allosteric effectors. $f(e)$ describes the dependency of the rate upon the enzyme concentration and can be taken to equal $V_{\max}$, while $g(\mathbf{X}, \mathbf{K})$ describes the interaction of the enzyme with the rest of metabolism through metabolite concentrations and the corresponding affinity constants. The important characteristic of the above equation is that f usually does not depend upon $\mathbf{X}$ and $\mathbf{K}$, and g does not depend upon e.

Considering the transition from one condition to another:

$$1 = \frac{\Delta \log f(e)}{\Delta \log J} \frac{\Delta \log g(\mathbf{X}, \mathbf{K})}{\Delta \log J} = \rho_h + \rho_m$$

ρ_h is called the "hierarchical regulation coefficient" (ter Kuile and Westerhoff 2001), since it quantifies the relative contribution of changes in enzyme capacity ($V_{\max}$) to the regulation of the enzyme's flux and depends on the complete gene-expression cascade of transcription, translation, posttranslational modification, and mRNA and protein degradation. ρ_m is the "metabolic regulation coefficient," quantifying the relative contribution of changes in the interaction of the enzyme with the rest of metabolism, to the regulation of the enzyme's flux. The two regulation coefficients sum up to one, implying that determination of one will yield the other automatically. In practice the hierarchical regulation coefficient is more readily determined, since $f(e)$ usually can be taken to equal $V_{\max}$, and the $V_{\max}$ as well as the flux J through the enzyme can be measured or estimated in many cases. Regulation analysis introduces the possibility of making unambiguous and quantitative descriptions of the regulation of fluxes through individual enzymes embedded in biochemical networks of any complexity, in response to any number or kind of simultaneous perturbations.

Regulation Analysis has been applied to understand the strong decrease of the glycolytic flux in the yeast *Saccharomyces cerevisiae* upon nutrient starvation (Rossell et al. 2005, 2006). It turned out that all possible combinations of metabolic and hierarchical regulation that one could possibly envisage, were indeed found *in vivo*. These include pure hierarchical regulation ($\rho_h = 1$), pure metabolic regulation ($\rho_m = 1$), cooperative regulation (both ρ_h and ρ_m positive and between 0 and 1) and antagonistic regulation (one regulation coefficient negative and the other one higher than 1) (Rossell et al. 2005, 2006).

A quantitative theory of regulation allows testing theories of regulation. Among the proposed mechanisms for metabolic flux changes, the two clearest hypotheses are: (i) modulation of single rate-limiting enzymes, and (ii) multi-site modulation, that is, the simultaneous and proportional modulation of all enzymes in the pathway, thus causing a change in flux while leaving metabolite concentrations unchanged (Fell and Thomas 1995). Although examples of multisite modulation exist (e.g. lipogenesis in mice, the urea cycle in rats, and photosynthesis in green plants (Fell and Thomas 1995)), they do not seem to represent a general mode of regulation. Upon a number of different perturbations the glycolytic pathway of bakers' yeast exhibited a complex regulation pattern in which each enzyme played a different role: leading the change, following or pulling back (Daran-Lapujade et al. 2006; Rossell et al. 2006).

When we know how much regulation is in gene expression and how much is in metabolism, it becomes interesting to investigate how this regulation is brought about at the molecular level. This makes Regulation Analysis a tool for making sense of metabolomics, proteomics, and transcriptomics data, or even for preceding such analyses in experiment design. Without such analysis one might get lost in large metabolomics patterns, but if one knows which enzymes are actually regulated at the metabolic level, one need to focus only on the metabolites that affect those enzymes.

Other metabolite levels may change, but these changes are apparently not regulating changes, but cancelled by others or unable to regulate enzyme rates. Similarly, if an enzyme is regulated (partly) by gene expression, we may investigate the quantitative importance of the various processes in the gene-expression cascade. The theory to do so has been developed recently and applied to yeast glycolysis (Daran-Lapujade et al. 2006).

A quantitative analysis such as Regulation Analysis requires highly accurate datasets. If we are to explain a 50% change of flux and the regulation is distributed over various processes, we need numbers with accuracy of less than 10% to identify processes that contribute 20% to the regulation. The analysis of *changes* between conditions is of course central in the study of regulation, but at the same time this makes it error-prone. Thanks to controlled cultivation and highly accurate assay techniques (Piper et al. 2002), transcriptome analysis is adequate to identify the most important regulatory phenomena, but there is a need for more accurate proteomics techniques (Daran-Lapujade et al. 2006).

3. THE SILICON CELL PROGRAMME

Systems Biology combines quantitative experimentation, computer modelling and theory (see Section 2) in order to come to an understanding of cell functioning. The Silicon Cell programme (Snoep et al. 2005, 2006) focuses on the construction of kinetic models of various aspects of cell functioning, such as metabolic, signal transduction and gene-expression models. Central to the Silicon Cell concept is that the models should be as realistic as possible. They are based on experimental data and are validated by comparing the model outcome to the behaviour of real cells. Although the ultimate ambition is to model the entire cell, at present models of subsystems are much more tractable. The idea is therefore to first construct models of subsystems, such as metabolic pathways (Schuster and Holzhutter 1995; Bakker et al. 1999; Teusink et al. 2000; Hoefnagel et al. 2002; Albert et al. 2005), signal transduction cascades (Schoeberl, Eichler-Jonsson et al. 2002; Bruggeman et al. 2005b) or gene-expression (Koster et al. 1988), study their properties, and later connect them to each other.

In view of the enormous effort that is made to make computer models, it is surprising that only recently attention has been paid to their preservation. Biological models were usually published as a set of differential equations. People who wanted to use the models, needed to re-programme them, often to find out that crucial data were missing or incorrect. This made it very difficult to revive a model without the help of the authors. A few years ago, however, Snoep and colleagues started to collect published models in an online database (JWS Online, www.jjj.bio.vu.nl (Snoep 2005)) that can be interrogated via the world-wide web. Today the database consists of some 50 models. Most of these models are realistic kinetic models of the Silicon Cell type, but the database also contains core models and demonstration models that serve to illustrate and investigate general principles and are very useful for teaching of systems biology.

A first attempt to link three models in the database, that is, a detailed glycolysis model of yeast (Teusink et al. 2000), a model of the branch to glycerol (Cronwright et al. 2002) and a model of glyoxylate production (Martins et al. 2001) was successful (Snoep et al. 2006). The behaviour of the glycolysis model was not so much altered by the addition of the two branches. In the combined model, the two branches, however, worked under other conditions than when they were considered in isolation. Their substrates and products were constant in the isolated models, while they were variables in the linked model. Of course various problems may occur when linking models systematically. For example, the models of subsystems may be valid for different experimental conditions or different cell types. Also they may overlap, in which case one has to choose which reactions to take from which model. Or there may be gaps between the models, which requires additional data. This emphasizes the need for standardization in systems biology. Several consortia are active in standardization. Computer scripts from different programmes can be interconverted via the Systems Biology Markup Language that is now supported by over a hundred software systems (Hucka et al. 2003) (http://sbml.org/). For example the Yeast Systems Biology Network (http://www.gmm.gu.se/YSBN/) is active in standardization and integration of Systems Biology activities of yeast, one of the major model organisms in the field. STRENDA (Standards for Reporting Enzymology Data) is active in the standarization of collecting and reporting enzyme data, which are vital for model construction (http://www.strenda.org).

4. ADVANTAGES OF A MODULAR APPROACH

4.1. Modules May Help Us Understand Cells Better

As we have seen above, a good strategy to make realistic kinetic models at a genome-wide scale is to start from smaller modules and to connect these later. There are other reasons to study the cell as if it consists of a number of modules. One is intelligibility. There is an interesting tension in offering quantitative explanations of cellular phenomena. As soon as the explanation becomes more quantitative it will necessarily contain more detail, describing reality more closely. There is a risk that we will miss the general principles by loosing ourselves in the study of details. To keep things understandable, we may temporarily treat parts that are not relevant for our research question at hand, as black boxes. Another reason to modularize models of the cell is that we may be interested in a certain process in the cell as it works *in the cellular context*, but we do not (yet) have the data to model the whole cell. Then we could treat the rest of the cell as a black box, but make sure that the compounds at the boundaries of the module of interest are described well. A final reason to modularize the cell is to determine the control exerted by certain processes in the cell (cf. Section 2.4), when it is not possible to influence these processes directly. In that case one might deduce the control from the elasticity coefficients, using the theorems of metabolic control analysis. Since these theorems only hold for the entire system, this can be done by dissecting the cell in modules of which the elasticity coefficients

can be determined experimentally. This is called modular (Schuster et al. 1993) or top-down control analysis.

An intuitive definition of a module would be a subnetwork inside the cell that performs a particular function relatively independently; a module is conceived as a functional unit. Then one could refer to this subnetwork rather than to its internal organization when explaining phenomena occurring at the (higher) cellular level. Examples of such kinds of modules are the ribosome and the degradosome, both macromolecular complexes with a well-defined function. More encompassing modules involving more proteins, involved in metabolism and signalling, and in wel defined pathways. This appears problematic, as sub-networks may be shared between functional modules. On the other hand, the definition of what constitutes a module may also follow other criteria: A more approximate explanation may require a more coarse-grained modularization of the system than a more detailed explanation. All in all, the defining of modules is not unambiguous. An interesting option is to define a module within a biochemical network such that it corresponds with a structural unit of organization, such as an organelle (mitochondrion, lysosome). Because organelles tend to have a limited number of transporters and receptors through which they interact with the rest of the cell, such a definition may indeed identify a functional unit of the network (see below). In hierarchical control analysis (Westerhoff et al. 1989; Kahn and Westerhoff 1991; Hofmeyr and Westerhoff 2001), module definition corresponded with the levels in the gene-expression hierarchy.

Another indication of the problems associated with modularization of complex systems is the small number of formal mathematical methods that allow one to simplify kinetic models. The existing methods are all based on time-scale separation in the system which allows for the decomposition of the system into a module composed of fast processes and one composed of slow processes. Then the fast processes can be considered in the absence of the slow processes. The slow processes are then considered with the fast processes either in steady state or thermodynamic equilibrium (Klonowski 1983; Segel and Slemrod 1989; Schuster and Schuster 1991; Kholodenko et al. 1998; Stiefenhofer 1998; Schneider and Wilhelm 2000). Two successful approaches to modularization of complex networks do not consider dynamics. One is purely structural while the other is applicable only to systems in steady state and concerns the analysis of control.

4.2. Examples of Successful Definitions of Modules

As the first example, we shall discuss two notions of modules often used in topological models of networks (Milo et al. 2002; Ravasz et al. 2002; Itzkovitz et al. 2003; Clauset et al. 2004; Newman 2006). The first notion is that of a "cluster" on the basis of the nodes within the cluster having more interactions among themselves than with nodes outside of the cluster. Clusters can be quantitatively defined on the basis of a cluster coefficient. *Escherichia coli* has been shown to be clustered in a hierarchical fashion (Ravasz et al. 2002). Another definition is that of a network motif (Milo et al. 2002; Shen-Orr et al. 2002; Kashtan et al. 2004; Yeger-Lotem et al. 2004). These authors

take a different strategy to define modules, or network motifs as they call them. A network motif is defined as a topological structure that appears with a higher frequency in a biological network than in a random network of the same size. Then the authors reason that this network topology has been selected for and should carry out some particular function contributing to fitness. They indeed found a number of such networks leading to a series of papers about their functional characterization (Mangan and Alon 2003; Mangan et al. 2003; Dekel et al. 2005; Kalir et al. 2005; Mangan et al. 2006). This is a promising approach but the many networks that do not appear significantly often, remain unaddressed. Do those networks then not carry out some function contributing to fitness? For instance, the incoherent feedforward loop network, for example, where X activates Y and inactivates Z and where Z is in addition activated by Y, does occur in *E. coli*'s transcription network, albeit not as frequent as the coherent feedforward loop network (where X activates Z directly rather than inactivating it), but does this mean that it fulfils no function inside *E. coli*? The answer is most certainly "no": it does perform a function – maybe one which is not required so abundantly or one which can also be fulfilled by another network structure occurring more frequently.

The second example of the use of modules derives from metabolic control analysis (Groen et al. 1982; Westerhoff and Van Dam 1987; Kahn and Westerhoff 1991; Schuster et al. 1993; Krab 1995; Brand 1996; van der Gugten and Westerhoff 1997; Ainscow and Brand 1999a, b; Krab et al. 2000; Krauss and Brand 2000; Hofmeyr and Westerhoff 2001; Bruggeman et al. 2002). This branch of metabolic control analysis is often referred to as hierarchical control analysis when the modules interact solely by regulatory influences (e.g. signal transduction cascades) and to modular control analysis if the modules are coupled by way of mass flow. Modular control analysis is sometimes referred to as top-down control analysis (Ainscow and Brand 1995; Brand 1996). In hierarchical control analysis the modules are often referred to as levels. In control analysis a level or a module is defined as a network that behaves the same in isolation of the network upon a parameter perturbation as it would behave in the network when the remainder of the network would be kept fixed upon perturbation of the same parameter. Mathematical analysis then shows that the only conditions for a module to behave sufficiently autonomously is that it should not share any conserved moieties with other modules (Schuster et al. 1993; Heinrich and Schuster 1996). The latter condition is important as it does limit the applicability of modularization to metabolism. On the other hand, there is a reward if modularization is possible: In modular control analysis, control coefficients can be defined for the module as a whole and the module can be considered a supra-enzyme in the entire network (Schuster et al. 1993; Ainscow and Brand 1995; Brand 1996; Ainscow and Brand 1999; Bruggeman et al. 2002). This leads to black-box modules and to an enormous simplification of the analysis of the control properties of the system under study (Schuster et al. 1993; Bruggeman et al. 2002). Importantly, the simplification may or may not correspond to what has been done intuitively for years, that is, discuss cell function in terms of interacting pathways, types of process (e.g. energy metabolism, carbon metabolism, transcription) and organelles.

4.3. Organelles as Natural Modules and Targets for Regulation

Indeed, compartmentation of the cell into various organelles that only communicate via molecules that are outside the organelle membrane or are transported across this membrane by well-defined carrier proteins, suggests a natural way to dissect the cell into modules. The notion that the functioning of the isolated organelle is relevant to the functioning of the whole cell is reflected for example by a long history of research on oxidative phosphorylation in isolated mitochondria, eventually leading to complete analysis of the control distribution (Groen et al. 1982; Ciapaite et al. 2005; Affourtit and Brand 2006) and the construction of detailed models (Demin et al. 1998; Demin et al. 2001).

Since organelles are often specialized to carry out a specific function, it makes sense to regulate the number of organelles depending on the needs of the cell. This is well known for the regulation of the number of mitochondria in yeast as a function of glucose concentration (Cho et al. 2001; Taylor et al. 2005) and in mammals as a function of thyroid hormone (Moraes 2001). Another clear example is the regulation of the number of alcohol oxidase containing peroxisomes in the yeast *Hansenula polymorpha*, depending on the presence of methanol in the medium (Leao-Helder et al. 2003).

In Section 2.3 we discussed the regulation of flux on an enzyme-by-enzyme basis. The overall regulation of flux was dissected in regulation of gene expression and metabolic regulation. However, the organelle abundance represents an additional level of regulation that can be incorporated formally in the theory of regulation analysis. If the organelle copy number per cell is included in the enzyme rate equation, we obtain:

$$v = V^*_{\max} \cdot \{\text{number of organelles per cell}\} \cdot g(\mathbf{X}, \mathbf{K})$$

in which v is again the *in vivo* rate of the enzyme, $V^*_{\max}$ is the maximum enzyme capacity *per organelle* (hence the *) and $g(\mathbf{X}, \mathbf{K})$ is again the metabolic part of the rate equation. Since at steady state the rate v equals the flux J, the comparison of the flux under two experimental conditions yields:

$$\frac{\Delta \log v}{\Delta \log J} = 1 = \frac{\Delta \log V_{\max}}{\Delta \log J} + \frac{\Delta \log\{\text{number of organelles per cell}\}}{\Delta \log J} + \frac{\Delta \log g(\mathbf{X}, \mathbf{K})}{\Delta \log J}$$
$$= \rho_h + \rho_{\text{organelle}} + \rho_m$$

in which $\rho_{\text{organelle}}$ now quantifies the importance of organelle abundance in the regulation of flux.

This is a new development for which we do not yet have experimental applications. On the other hand, *T. brucei* with its glycosomes, the concentration of which changes when the parasite shifts between its various stages in its life cycle, seems a most interesting test case.

5. CONCLUDING REMARKS

From the above it is clear that Systems Biology is on its way to reconstruct the living cell and subsequently living organisms in terms of their molecular components. There is much activity at the level of intracellular biochemical reaction, signalling and gene-expression networks. In the context of this book, it is most important to note that there is comparatively little activity in Systems Biology that analyzes the living cell in terms of dynamically interacting structural components. The reason is that Life is fundamentally a non-equilibrium condition that is maintained by a number of physical chemical processes. Much of the existing paradigm of Systems Biology takes the spatial position of all these processes either as homogeneous within the cell, or as well-defined by the position of the enzyme or transporter catalyzing the molecular process. Systems Biology has not yet much been emphasizing that the spatial organization of the molecular components, depends strongly on the presence of intracellular structures.

Making the link from structural organization is sometimes considered trivial. It is not, however. The intracellular structures themselves are not constants but depend on the effect of a multitude of cooperating non-equilibrium processes. This is clear for the supercoiling state of prokaryotic DNA (Snoep et al. 2002), but also for the complex structure of chromatin in higher eukaryotes.

Intriguingly, intracellular structures are the consequence of a number of processes that happen more or less instantaneously (and may be considered Markovian technically), but also of a substantial number of other processes that may be inherited from parental cells. Many of the ribosomes in a cell may in fact have been synthesized in the maternal cell, hence in a completely different context. Most of the plasma membrane of a cell originates in the maternal cell. Consequently, also much of the intracellular dynamic networking of cells may be much more hysteretic than assumed by standard systems biology. Another exciting aspect is the problem of coupling the generation and maintenance of structures that are asymmetrical in space to biochemical reactions that could just as well have been symmetrical. The answer to this issue is known to be the spatial organization of the enzymes, leading to yet another example of the circular or spiralling causality so characteristic of Systems Biology.

The Systems Biology of dynamic intracellular structures and compartmentation has an exciting and bright future. We can only hope that this chapter can help start the new activities.

REFERENCES

Acerenza, L., Sauro, H.M. and Kacser, H. (1989) Control analysis of time-dependent metabolic systems. J. Theor. Biol. *137*, 423–444.

Affourtit, C. and Brand, M.D. (2006) Stronger control of ATP/ADP by proton leak in pancreatic beta-cells than skeletal muscle mitochondria. Biochem. J. *393*, 151–159.

Ainscow, E.K. and Brand, M.D. (1995) Top-down control analysis of systems with more than one common intermediate. Eur. J. Biochem. *231*, 579–586.

Ainscow, E.K. and Brand, M.D. (1999a) Internal regulation of ATP turnover, glycolysis and oxidative phosphorylation in rat hepatocytes. Eur. J. Biochem. *266*, 737–749.

Ainscow, E.K. and Brand, M.D. (1999b) Top-down control analysis of ATP turnover, glycolysis and oxidative phosphorylation in rat hepatocytes. Eur. J. Biochem. *263*, 671–685.

Alberghina, L. and Westerhoff, H.V. eds. (2005) Systems Biology, Definitions and Perspectives (Topics in Current Genetics), (Berlin: Springer-Verlag and Heidelberg: GmbH & Co. K.).

Albert, R. and Barabasi, A.L. (2002) Statistical mechanics of complex networks. Rev. Mod. Phys. *74*, 47–97.

Albert, M.A., Haanstra, J.R, Hannaert, V., Van Roy, J., Opperdoes, F.R., Bakker, B.M. and Michels, P.A. (2005) Experimental and in silico analyses of glycolytic flux control in bloodstream form *Trypanosoma brucei*. J. Biol. Chem. *280*, 28306–28315.

Allen, T.E. and Palsson, B.O. (2003) Sequence-based analysis of metabolic demands for protein synthesis in prokaryotes. J. Theor. Biol. *220*, 1–18.

Bakker, B.M., Michels, P.A.M., Opperdoes, F.R. and Westerhoff, H.V. (1997) Glycolysis in bloodstream form *Trypanosoma brucei* can be understood in terms of the kinetics of the glycolytic enzymes. J. Biol. Chem. *272*, 3207–3215.

Bakker, B.M., Walsh, M.C., ter Kuile, B.H., Mensonides, F.I., Michels, P.A., Opperdoes, F.R. and Westerhoff, H.V. (1999) Contribution of glucose transport to the control of the glycolytic flux in *Trypanosoma brucei*. Proc. Natl Acad. Sci. U.S.A. *96*, 10098–10103.

Brand, M.D. (1996) Top down metabolic control analysis. J. Theor. Biol. *182*, 351–360.

Brenner, S. (1999) Theoretical biology in the third millennium. Philos. Trans. R. Soc. Lond., B, Biol. Sci. *354*, 1963–1965.

Brown, G.C. and Kholodenko, B.N. (1999) Spatial gradients of cellular phospho-proteins. FEBS Lett. *457*, 452–454.

Bruggeman, F.J., Bakker, B.M., Hornberg, J.J. and Westerhoff, H.V. (2005a) Introduction to computational models of biochemical reaction networks. In: Computational Systems Biology. A. Kriete and R. Eils, eds. (London, UK: Elsevier Academic Press), pp. 127–148.

Bruggeman, F.J., Boogerd, F.C. and Westerhoff, H.V. (2005b) The multifarious short-term regulation of ammonium assimilation of *Escherichia coli*, dissection using an in silico replica. Febs. J. *272*, 1965–1985.

Bruggeman, F.J., Westerhoff, H.V., Hoek, J.B. and Kholodenko, B.N. (2002) Modular response analysis of cellular regulatory networks. J. Theor. Biol. *218*, 507–520.

Cho, J.H., Lee, Y.K. and Chae, C.B. (2001) The modulation of the biological activities of mitochondrial histone Abf2p by yeast PKA and its possible role in the regulation of mitochondrial DNA content during glucose repression. Biochim. Biophys. Acta *1522*, 175–186.

Ciapaite, J., Van Eikenhorst, G., Bakker, S.J.L., Diamant, M., Heine, R.J., Wagner, M.J., Westerhoff, H.V. and Krab, K. (2005) Modular kinetic analysis of the adenine nucleotide translocator-mediated effects of palmitoyl-CoA on the oxidative phosphorylation in isolated rat liver mitochondria. Diabetes *54*, 944–951.

Clauset, A., Newman, M.E. and Moore, C. (2004) Finding community structure in very large networks. Phys. Rev. E. Stat. Nonlin. Soft Matter Phys. *70*, 066111.

Cleland, W.W. (1963a). The kinetics of enzyme-catalyzed reactions with two or more substrates or products. I. Nomenclature and rate equations. Biochim. Biophys. Acta *67*, 104–137.

Cleland, W.W. (1963b). The kinetics of enzyme-catalyzed reactions with two or more substrates or products. II. Inhibition: nomenclature and theory. Biochim. Biophys. Acta *67*, 173–187.

Cleland, W.W. (1963c). The kinetics of enzyme-catalyzed reactions with two or more substrates or products. III. Prediction of initial velocity and inhibition patterns by inspection. Biochim. Biophys. Acta *67*, 188–196.

Cronwright, G.R., Rohwer, J.M. and Prior, B.A. (2002) Metabolic Control Analysis of Glycerol Synthesis in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. *68*, 4448–4456.

Daran-Lapujade, P., Rossell, S.L., Van Gulik, W., Luttik, M.A.H., De Groot, M., Slijper, M., Heck, A.J.R., Daran, J.M., De Winde, J.H., Westerhoff, H.V. et al. (2006) Manuscript in preparation.

Davies, S.E. and Brindle, K.M. (1992) Effects of overexpression of phosphofructokinase on glycolysis in the yeast *Saccharomyces cerevisiae*. Biochemistry *31*, 4729–4735.

Dekel, E., Mangan, S. and Alon, U. (2005) Environmental selection of the feed-forward loop circuit in gene-regulation networks. Phys. Biol. *2*, 81–88.

Delgado, J. and Liao, J.C. (1995) Control of metabolic pathways by time-scale separation. Biosystems *36*, 55–70.

Demin, O.V., Gorianin, I.I., Kholodenko, B.N. and Westerhoff, H.V. (2001) Kinetic modeling of energy metabolism and generation of active forms of oxygen in hepatocyte mitochondria. Mol. Biol. (Mosk.) *35*, 1095–1104.

Demin, O.V., Westerhoff, H.V. and Kholodenko, B.N. (1998) Mathematical modelling of superoxide generation with the bc1 complex of mitochondria. Biochemistry (Mosc.) *63*, 634–649.

Easterby, J.S. (1990) Integration of temporal analysis and control analysis of metabolic systems. Biochem. J. *269*, 255–259.

Edwards, J.S., Covert, M. and Palsson, B. (2002) Metabolic modelling of microbes: the flux-balance approach. Environ. Microbiol. *4*, 133–140.

Famili, I. and Palsson, B.O. (2003) The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. Biophys. J. *85*, 16–26.

Fell, D.A. (1997) Understanding the Control of Metabolism. First edn. (London and Miami: Portland Press).

Fell, D.A. and Sauro, H.M. (1985) Metabolic control and its analysis. Additional relationships between elasticities and control coefficients. Eur. J. Biochem. *148*, 555–561.

Fell, D.A. and Sauro, H.M. (1990) Metabolic control analysis. The effects of high enzyme concentrations. Eur. J. Biochem. *192*, 183–187.

Fell, D.A. and Thomas, S. (1995) Physiological control of metabolic flux: the requirement for multisite modulation. Biochem. J. *311*, 35–39.

Fell, D.A. and Wagner, A. (2000) The small world of metabolism. Nat. Biotechnol. *18*, 1121–1122.

Flint, H.J., Porteous, D.J. and Kacser, H. (1980) Control of the flux in the arginine pathway of *Neurospora crassa*. The flux from citrulline to arginine. Biochem. J. *190*, 1–15.

Flint, H.J., Tateson, R.W., Barthelmess, I.B., Porteous, D.J., Donachie, W.D. and Kacser, H. (1981) Control of the flux in the arginine pathway of *Neurospora crassa*. Modulations of enzyme activity and concentration. Biochem. J. *200*, 231–246.

Groen, A.K., Wanders, R.J., Westerhoff, H.V., van der Meer, R. and Tager, J.M. (1982) Quantification of the contribution of various steps to the control of mitochondrial respiration. J. Biol. Chem. *257*, 2754–2757.

Heinrich, R. and Rapoport, T.A. (1974) A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. Eur. J. Biochem. *42*, 89–95.

Heinrich, R. and Reder, C. (1991) Metabolic control analysis of relaxation processes. J. Theor. Biol. *151*, 57–61.

Heinrich, R. and Schuster, S. (1996) The Regulation of Cellular Systems. First edn. (New York: Chapman & Hall).

Hess, E.L. (1970) Origins of molecular biology. Science *168*, 664–669.

Hoefnagel, M.H.N., Starrenburg, M.J.C., Martens, D.E., Hugenholtz, J., Kleerebezem, M., Van Swam, I.I., Bongers, R., Westerhoff, H.V. and Snoep, J.L. (2002) Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. Microbiology *148*, 1003–1013.

Hofmeyr, J.H. (1995) Metabolic regulation, a control analytic perspective. J. Bioenerg. Biomembr. *27*, 479–490.

Hofmeyr, J.H., Cornish-Bowden, A. and Rohwer, J.M. (1993) Taking enzyme kinetics out of control; putting control into regulation. Eur. J. Biochem. *212*, 833–837.

Hofmeyr, J.H., Kacser, H. and van der Merwe, K.J. (1986) Metabolic control analysis of moiety-conserved cycles. Eur. J. Biochem. *155*, 631–641.

Hofmeyr, J.H. and Westerhoff, H.V. (2001) Building the cellular puzzle: control in multi-level reaction networks. J. Theor. Biol. *208*, 261–285.

Hood, L. (2003) Systems biology: integrating technology, biology, and computation. Mech. Ageing Dev. *124*, 9–16.

Hornberg, J.J., Bruggeman, F.J., Binder, B., Geest, C.R., de Vaate, A.J., Lankelma, J., Heinrich, R. and Westerhoff, H.V. (2005) Principles behind the multifarious control of signal transduction. ERK phosphorylation and kinase/phosphatase control. FEBS J. *272*, 244–258.

Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D. and Cornish-Bowden, A. et al. (2003) The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models. Bioinformatics *19*, 524–531.

Ideker, T., Galitski, T. and Hood, L. (2001a) A new approach to decoding life: systems biology. Annu. Rev. Genomics. Hum. Genet. *2*, 343–372.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R. and Hood, L. (2001b) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science *292*, 929–934.

Imielinski, M., Belta, C., Rubin, H. and Halasz, A. (2006) Systematic analysis of conservation relations in *Escherichia coli* genome-scale metabolic network reveals novel growth media. Biophys. J. *90*, 2659–2672.

Ingalls, B. (2004) Autonomously oscillating biochemical systems: parametric sensitivity of extrema and period. Syst. Biol. *1*, 62–70.

Ingalls, B.P. and Sauro, H.M. (2003) Sensitivity analysis of stoichiometric networks: an extension of metabolic control analysis to non-steady state trajectories. J. Theor. Biol. *222*, 23–36.

Itzkovitz, S., Milo, R., Kashtan, N., Ziv, G. and Alon, U. (2003) Subgraphs in random networks. Phys. Rev. E. Stat. Nonlin. Soft Matter Phys. *68*, 026127.

Jensen, P.R., Van Der Weijden, C.C., Jensen, L.B., Westerhoff, H.V. and Snoep, J.L. (1999) Extensive regulation compromises the extent to which DNA gyrase controls DNA supercoiling and growth rate of *Escherichia coli*. Eur. J. Biochem. *266*, 865–877.

Jeong, H., Mason, S.P., Barabasi, A.L. and Oltvai, Z.N. (2001) Lethality and centrality in protein networks. Nature *411*, 41–42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabasi, A.L. (2000) The large-scale organization of metabolic networks. Nature *407*, 651–654.

Kacser, H. and Burns, J.A. (1973) The control of flux. Symp. Soc. Exp. Biol. *27*, 65–104.

Kacser, H., Sauro, H.M. and Acerenza, L. (1990) Enzyme-enzyme interactions and control analysis. 1. The case of non-additivity: monomer-oligomer associations. Eur. J. Biochem. *187*, 481–491.

Kahn, D. and Westerhoff, H.V. (1991) Control theory of regulatory cascades. J. Theor. Biol. *153*, 255–285.

Kahn, D. and Westerhoff, H.V. (1993) Regulation and homeostasis in metabolic control theory: interplay between fluctuations of variables and parameter changes. In: Modern Trends in BioThermoKinetics. S. Schuster, M. Rigoulet, R. Ouhabi, J.P. Mazat, ed. (New York: Plenum Press), pp. 199–204.

Kalir, S., Mangan, S. and Alon, U. (2005) A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. Mol. Syst. Biol. *1*, 0006.

Kashtan, N., Itzkovitz, S., Milo, R. and Alon, U. (2004) Topological generalizations of network motifs. Phys. Rev. E. Stat. Nonlin. Soft Matter Phys. *70*, 031909.

Kholodenko, B.N., Cascante, M. and Westerhoff, H.V. (1994a) Control theory of metabolic channelling. Mol. Cell. Biochem. *133-134*, 313–331.

Kholodenko, B.N., Demin, O.V. and Westerhoff, H.V. (1996) The metabolic control theory of biochemical oscillating systems. 1. Definitions of the quantitative characteristics and their simplest properties. Biochemistry Mosc. *61*, 423–434.

Kholodenko, B.N., Demin, O.V. and Westerhoff, H.V. (1997a) Control analysis of periodic phenomena in biological systems. J. Phys. Chem. B *101*, 2070–2081.

Kholodenko, B.N., Demin, O.V., Moehren, G. and Hoek, J.B. (1999) Quantification of short term signaling by the epidermal growth factor receptor. J. Biol. Chem. *274*, 30169–30181.

Kholodenko, B.N., Hoek, J.B., Westerhoff, H.V. and Brown, G.C. (1997b) Quantification of information transfer via cellular signal transduction pathways. FEBS Lett. *414*, 430–434.

Kholodenko, B.N., Molenaar, D., Schuster, S., Heinrich, R. and Westerhoff, H.V. (1995a) Defining control coefficients in non-ideal metabolic pathways. Biophys. Chem. *56*, 215–226.

Kholodenko, B.N., Sauro, H.M. and Westerhoff, H.V. (1994b) Control by enzymes, coenzymes and conserved moieties. A generalisation of the connectivity theorem of metabolic control analysis. Eur. J. Biochem. *225*, 179–186.

Kholodenko, B.N., Sauro, H.M., Westerhoff, H.V. and Cascante, M. (1995b) Coenzyme cycles and metabolic control analysis: the determination of the elasticity coefficients from the generalised connectivity theorem. Biochem. Mol. Biol. Int. *35*, 615–625.

Kholodenko, B.N., Schuster, S., Garcia, J., Westerhoff, H.V. and Cascante, M. (1998) Control analysis of metabolic systems involving quasi-equilibrium reactions. Biochim. Biophys. Acta *1379*, 337–352.

Kholodenko, B.N. and Westerhoff, H.V. (1994) Control theory of one enzyme. Biochim. Biophys. Acta *1208*, 294–305.

Kitano, H. (2002) Systems biology: a brief overview. Science *295*, 1662–1664.

Klonowski, W. (1983) Simplifying principles for chemical and enzyme reaction kinetics. Biophys. Chem. *18*, 73–87.

Koster, J.G., Destrée, O.H.J. and Westerhoff, H.V. (1988) Kinetics of histone gene expression during early development of *Xenopus laevis*. J. Theor. Biol. *135*, 139–167.

Krab, K. (1995) Kinetic and regulatory aspects of the function of the alternative oxidase in plant respiration. J. Bioenerg. Biomembr. *27*, 387–396.

Krab, K., Wagner, M.J., Wagner, A.M. and Moller, I.M. (2000) Identification of the site where the electron transfer chain of plant mitochondria is stimulated by electrostatic charge screening. Eur. J. Biochem. *267*, 869–876.

Krauss, S. and Brand, M.D. (2000) Quantitation of signal transduction. FASEB J. *14*, 2581–2588.

Kriete, A. and Eils, R. eds. (2005) Computational Systems Biology. (London, UK: Elsevier Academic press).

Leao-Helder, A.N., Krikken, A.M., Van der Klei, I.J., Kiel, J.A. and Veenhuis, M. (2003) Transcriptional down-regulation of peroxisome numbers affects selective peroxisome degradation in *Hansenula polymorpha*. J. Biol. Chem. *278*, 40749–40756.

Mahadevan, R. and Schilling, C.H. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. *5*, 264–276.

Mangan, S. and Alon, U. (2003) Structure and function of the feed-forward loop network motif. Proc. Natl Acad. Sci. U.S.A. *100*, 11980–11985.

Mangan, S., Itzkovitz, S., Zaslaver, A. and Alon, U. (2006) The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. J. Mol. Biol. *356*, 1073–1081.

Mangan, S., Zaslaver, A. and Alon, U. (2003) The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. J. Mol. Biol. *334*, 197–204.

Martins, A.M., Mendes, P., Cordeiro, C. and Freire, A.P. (2001) In situ kinetic analysis of glyoxalase I and glyoxalase II in *Saccharomyces cerevisiae*. Eur. J. Biochem. *268*, 3930–3936.

Melendezhevia, E., Torres, N.V., Sicilia, J. and Kacser, H. (1990) Control analysis of transition times in metabolic systems. Biochem. J. *265*, 195–202.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002) Network motifs: simple building blocks of complex networks. Science *298*, 824–827.

Moraes, C.T. (2001) What regulates mitochondrial DNA copy number in animal cells? Trends Genet. *17*, 199–205.

Newman, M.E. (2006) From the Cover: modularity and community structure in networks. Proc. Natl Acad. Sci. U.S.A. *103*, 8577–8582.

Newman, M.E.J. (2003) The structure and function of complex networks. SIAM Rev. *45*, 167–256.

Nikolaev, E.V., Burgard, A.P. and Maranas, C.D. (2005) Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. Biophys. J. *88*, 37–49.

Papin, J.A., Hunter, T., Palsson, B.O. and Subramaniam, S. (2005) Reconstruction of cellular signalling networks and analysis of their properties. Nat. Rev. Mol. Cell. Biol. *6*, 99–111.

Papin, J.A. and Palsson, B.O. (2004a). The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis. Biophys. J. *87*, 37–46.

Papin, J.A. and Palsson, B.O. (2004b). Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. J. Theor. Biol. *227*, 283–297.

Papin, J.A., Stelling, J., Price, N.D., Klamt, S., Schuster, S. and Palsson, B.O. (2004) Comparison of network-based pathway analysis methods. Trends Biotechnol. *22*, 400–405.

Peletier, M.A., Westerhoff, H.V. and Kholodenko, B.N. (2003) Control of spatially heterogeneous and time-varying cellular reaction networks: a new summation law. J. Theor. Biol. *225*, 477–487.

Pereira-Leal, J.B., Enright, A.J. and Ouzounis, C.A. (2004) Detection of functional modules from protein interaction networks. Proteins *54*, 49–57.

Piper, M.D., Daran-Lapujade, P., Bro, C., Regenberg, B., Knudsen, S., Nielsen, J. and Pronk, J.T. (2002) Reproducibility of oligonucleotide microarray transcriptome analyses. An interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*. J. Biol. Chem. *277*, 37001–37008.

Price, N.D., Reed, J.L. and Palsson, B.O. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat. Rev. Microbiol. *2*, 886–897.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabasi, A.L. (2002) Hierarchical organization of modularity in metabolic networks. Science *297*, 1551–1555.

Reder, C. (1988) Metabolic control theory: a structural approach. J. Theor. Biol. *135*, 175–201.

Reed, J.L. and Palsson, B.O. (2004) Genome-scale in silico models of *E.coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. Genome Res. *14*, 1797–1805.

Rohwer, J.M., Meadow, N.D., Roseman, S., Westerhoff, H.V. and Postma, P.W. (2000) Understanding glucose transport by the bacterial phosphoenolpyruvate:glycose phosphotransferase system on the basis of kinetic measurements *in vitro*. J. Biol. Chem. *275*, 34909–34921.

Rossell, S., van der Weijden, C.C., Kruckeberg, A.L., Bakker, B.M. and Westerhoff, H.V. (2005) Hierarchical and metabolic regulation of glucose influx in starved *Saccharomyces cerevisiae*. FEMS Yeast Res. *5*, 611–619.

Rossell, S., van der Weijden, C.C., Lindenbergh, A., van Tuijl, A., Francke, C., Bakker, B.M. and Westerhoff, H.V. (2006) Unraveling the complexity of flux regulation: A new method demonstrated for nutrient starvation in *Saccharomyces cerevisiae*. Proc. Natl Acad. Sci. U.S.A. *103*, 2166–2171.

Ruijter, G.J.G., Postma, P.W. and Van Dam, K. (1991) Control of glucose metabolism by enzyme II^{Glc} of the phosphoenolpyruvate-dependent phosphotransferase system in *Escherichia coli*. J. Bacteriol. *173*, 6184–6191.

Sauro, H.M. (1994) Moiety-conserved cycles and metabolic control analysis: problems in sequestration and metabolic channelling. Biosystems *33*, 55–67.

Sauro, H.M. and Ingalls, B. (2004) Conservation analysis in biochemical networks: computational issues for software writers. Biophys. Chem. *109*, 1–15.

Sauro, H.M. and Kacser, H. (1990) Enzyme-enzyme interactions and control analysis. 2. The case of non-independence: heterologous associations. Eur. J. Biochem. *187*, 493–500.

Sauro, H.M., Small, J.R. and Fell, D.A. (1987) Metabolic control and its analysis. Extensions to the theory and matrix method. Eur. J. Biochem. *165*, 215–221.

Schaaff, I., Heinisch, J. and Zimmermann, F.K. (1989) Overproduction of glycolytic enzymes in yeast. Yeast *5*, 285–290.

Schilling, C.H., Letscher, D. and Palsson, B.O. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J. Theor. Biol. *203*, 229–248.

Schilling, C.H., Schuster, S., Palsson, B.O. and Heinrich, R. (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol. Prog. *15*, 296–303.

Schneider, K.R. and Wilhelm, T. (2000) Model reduction by extended quasi-steady-state approximation. J. Math. Biol. *40*, 443–450.

Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D. and Muller, G. (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat. Biotechnol. *20*, 370–375.

Schuster, S. (1996) Control analysis in terms of generalized variables characterizing metabolic systems. J. Theor. Biol. *182*, 259–268.

Schuster, S., Dandekar, T. and Fell, D.A. (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol. *17*, 53–60.

Schuster, S., Fell, D.A. and Dandekar, T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat. Biotechnol. *18*, 326–332.

Schuster, R. and Holzhutter, H.G. (1995) Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations. Application to enzyme deficiencies of red blood cells. Eur. J. Biochem. *229*, 403–418.

Schuster, S., Kahn, D. and Westerhoff, H.V. (1993) Modular analysis of the control of complex metabolic pathways. Biophys. Chem. *48*, 1–17.

Schuster, R. and Schuster, S. (1991) Relationships between modal-analysis and rapid-equilibrium approximation in the modeling of biochemical networks. Syst. Anal. Model. Simul. *8*, 623–633.

Segel, I.H. (1993) Enzyme kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-state Enzyme Systems. (New York: John Wiley & Sons, Inc.).

Segel, L.A. and Slemrod, M. (1989) The quasi-steady-state assumption - a case-study in perturbation. Siam Rev. *31*, 446–477.

Shen-Orr, S.S., Milo, R., Mangan, S. and Alon, U. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat. Genet. *31*, 64–68.

Snoep, J.L. (2005) The Silicon Cell initiative: working towards a detailed kinetic description at the cellular level. Curr. Opin. Biotechnol. *16*, 336–343.

Snoep, J.L., Arfman, N., Yomano, L.P., Westerhoff, H.V., Conway, T. and Ingram, L.O. (1996) Control of glycolytic flux in *Zymomonas mobilis* by glucose-6-phosphate dehydrogenase activity. Biotechnol. Bioeng. *51*, 190–197.

Snoep, J.L., Bruggeman, F., Olivier, B.G. and Westerhoff, H.V. (2006) Towards building the silicon cell: a modular approach. BioSystems *83*, 207–216.

Snoep, J.L., van der Weijden, C.C., Andersen, H.W., Westerhoff, H.V. and Jensen, P.R. (2002) DNA supercoiling in *Escherichia coli* is under tight and subtle homeostatic control, involving gene-expression and metabolic regulation of both topoisomerase I and DNA gyrase. Eur. J. Biochem. *269*, 1662–1669.

Spirin, V. and Mirny, L.A. (2003) Protein complexes and functional modules in molecular networks. Proc. Natl Acad. Sci. U.S.A. *100*, 12123–12128.

Stent, G.S. (1968) That was the molecular biology that was. Science *160*, 390–395.

Stiefenhofer, M. (1998) Quasi-steady-state approximation for chemical reaction networks. J. Math. Biol. *36*, 593–609.

Taylor, S.D., Zhang, H., Eaton, J.S., Rodeheffer, M.S., Lebedeva, M.A., O'rourke, T.W., Siede, W. and Shadel, G.S. (2005) The conserved Mec1/Rad53 nuclear checkpoint pathway regulates mitochondrial DNA copy number in *Saccharomyces cerevisiae*. Mol. Biol. Cell. *16*, 3010–3018.

ter Kuile, B.H. and Westerhoff, H.V. (2001) Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. FEBS Lett. *500*, 169–171.

Teusink, B., Passarge, J., Reijenga, C.A., Esgalhado, E., van der Weijden, C.C., Schepper, M., Walsh, M.C., Bakker, B.M., van Dam, K., Westerhoff, H.V. and Snoep, J.L. (2000) Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. Eur. J. Biochem. *267*, 5313–5329.

van der Gugten, A.A. and Westerhoff, H.V. (1997) Internal regulation of a modular system: the different faces of internal control. Biosystems *44*, 79–106.

Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of "small-world" networks. Nature *393*, 440–442.

Westerhoff, H.V. and Chen, Y.D. (1984) How do enzyme activities control metabolite concentrations? An additional theorem in the theory of metabolic control. Eur. J. Biochem. *142*, 425–430.

Westerhoff, H.V. and Kell, D.B. (1996) What bio technologists knew all along...? J. Theor. Biol. *182*, 411–420.

Westerhoff, H.V., Koster, J.G., van Workum, M. and Rudd, K.E. (1989) On the control of gene expression. In: Control of Metabolic Processes. A. Cornish-Bowden, M.L. Cardenas, eds. (New York: Plenum Press), pp. 399–413, NATO ASI Series.

Westerhoff, H.V. and Palsson, B.O. (2004) The evolution of molecular biology into systems biology. Nat. Biotechnol. *22*, 1249–1252.

Westerhoff, H.V., Plomp, P.J., Groen, A.K. and Wanders, R.J. (1987) Thermodynamics of the control of metabolism. Cell Biophys. *11*, 239–267.

Westerhoff, H.V. and Van Dam, K. (1987) Thermodynamics and Control of Biological Free-energy Transduction. (Amsterdam: Elsevier Science Publishers B.V. (Biomedical Division)).

Wiechert, W. (2002) An introduction to 13C metabolic flux analysis. Genet. Eng. (NY) *24*, 215–238.

Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R.Y., Alon, U. and Margalit, H. (2004) Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. Proc. Natl Acad. Sci. U.S.A. *101*, 5934–5939.

Corresponding author: Dr. Barbara Bakker (barbara.bakker@falw.vu.nl)