Systems Biology towards life in silico: mathematics of the control of living cells.

Westerhoff, H.V.; Kolodkin, A.N.; Conradie, R.; Wilkinson, S.J.; Bruggeman, F.J.; Krab, K.; Schuppen, J.H.; Hardin, H.M.; Bakker, B.M.; Mone, M.J.; Rybakova, E.N.; Eijken, M.; van Leeuwen, J.P.; Snoep, J.L.

published in
Journal of Mathematical Biology
2009

DOI (link to publisher)
10.1007/s00285-008-0160-8

document version
Publisher's PDF, also known as Version of record

J. Math. Biol. (2009) 58:7–34
DOI 10.1007/s00285-008-0160-8

Mathematical Biology

Systems biology towards life in silico: mathematics of the control of living cells

Hans V. Westerhoff · Alexey Kolodkin · Riaan Conradie · Stephen J. Wilkinson · Frank J. Bruggeman · Klaas Krab · Jan H. van Schuppen · Hanna Hardin · Barbara M. Bakker · Martijn J. Moné · Katja N. Rybakova · Marco Eijken · Hans J. P. van Leeuwen · Jacky L. Snoep

Received: 26 October 2007 / Revised: 11 January 2008 / Published online: 16 February 2008

Abstract Systems Biology is the science that aims to understand how biological function absent from macromolecules in isolation, arises when they are components of their system. Dedicated to the memory of Reinhart Heinrich, this paper discusses the origin and evolution of the new part of systems biology that relates to met-

H. V. Westerhoff (✉)
Molecular Cell Physiology, Institute of Molecular Cell Biology,
Netherlands Institute for Systems Biology, Faculty of Earth and Life Sciences,
Vrije Universiteit, De Boelelaan 1085,
1081 HV Amsterdam, The Netherlands
e-mail: hans.westerhoff@manchester.ac.uk

A. Kolodkin · F. J. Bruggeman · K. Krab · J. H. van Schuppen ·
H. Hardin · B. M. Bakker · M. J. Moné · K. N. Rybakova · J. L. Snoep
Molecular Cell Physiology, Institute of Molecular Cell Biology, BioCentrum Amsterdam,
Faculty for Earth and Life Sciences, Vrije Universiteit, De Boelelaan 1085,
1081 HV Amsterdam, The Netherlands

H. V. Westerhoff · S. J. Wilkinson · J. L. Snoep
Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre,
School of Chemical Engineering and Analytical Science, The University of Manchester,
131 Princess Street, Manchester M1 7ND, UK

R. Conradie · J. L. Snoep
Department of Biochemistry, Stellenbosch University,
Private Bag X1, Matieland 7602, South Africa

F. J. Bruggeman · J. H. van Schuppen · H. Hardin
CWI, 1090 GB Amsterdam, The Netherlands

M. Eijken · H. J. P. van Leeuwen
Department of Internal Medicine, Erasmus Medical Centre,
3000 DR Rotterdam, The Netherlands

abolic and signal-transduction pathways and extends mathematical biology so as to address postgenomic experimental reality. Various approaches to modeling the dynamics generated by metabolic and signal-transduction pathways are compared. The silicon cell approach aims to describe the intracellular network of interest precisely, by numerically integrating the precise rate equations that characterize the ways macromolecules' interact with each other. The non-equilibrium thermodynamic or 'lin–log' approach approximates the enzyme rate equations in terms of linear functions of the logarithms of the concentrations. Biochemical Systems Analysis approximates in terms of power laws. Importantly all these approaches link system behavior to molecular interaction properties. The latter two do this less precisely but enable analytical solutions. By limiting the questions asked, to optimal flux patterns, or to control of fluxes and concentrations around the (patho)physiological state, Flux Balance Analysis and Metabolic/Hierarchical Control Analysis again enable analytical solutions. Both the silicon cell approach and Metabolic/Hierarchical Control Analysis are able to highlight where and how system function derives from molecular interactions. The latter approach has also discovered a set of fundamental principles underlying the control of biological systems. The new law that relates concentration control to control by time is illustrated for an important signal transduction pathway, i.e. nuclear hormone receptor signaling such as relevant to bone formation. It is envisaged that there is much more Mathematical Biology to be discovered in the area between molecules and Life.

Keywords Systems biology · Systems behaviour · Modeling · Silicon cell · Metabolic control analysis · Flux balance analysis · Glucocorticoid receptors · Signal transduction

Mathematics Subject Classification (2000) 93A30

Contents

1 Introduction

In the group of Sam Rapoport in Berlin, in the late 1960s, there was great interest in red blood cells [35]. One practical motivation was the issue of storage of blood and of erythrocytes to be used for transfusion. Scientifically this system was attractive because so much experimental material was readily available (dated blood) and because of its relative simplicity. Mitochondria are absent from human erythrocytes such that their ATP is produced only glycolytically. In addition the absence of nuclei causes the pathways to be present at relatively constant levels, not to be subject to induction and adaptation, although degradation of enzymes during the life cycle of the red blood cell might be important.

The availability of the experimental material enabled the Rapoport group to characterize the glycolytic enzymes kinetically. Coming from a background of theoretical physics but with his strong interest in much of Life, Reinhart Heinrich, with the help of the two Rapoports, put together the experimental information into a kinetic model of glycolysis. The extensive paper in Progress in Biophysics and Molecular Biology by Heinrich et al. [20] has been seminal to much work to follow. It led to a persistent research program around Heinrich and the Rapoports, engaging others like Jacobasch, Holzhütter and the Schusters. It has also provided much of the intellectual basis of the silicon cell concept (http://www.siliconcell.net), which maintains that by specifying the properties of the enzymes in a metabolic pathway also in terms of their interactions with metabolites, one can reconstruct in the computer the behavior of the pathway.

This work has led to a substantial development in Mathematical Biology, focusing on biochemical reaction networks and presently contributing substantially to Systems Biology. In this paper, which we dedicate to the late Reinhart Heinrich, we give an overview of the corresponding field and close with a new extension to the dynamic analysis of signal transduction in time.

2 Networks in cell biology

2.1 Systems biology

Systems Biology is the science that aims to understand how biological function that is absent from macromolecules in isolation, arises when they are components in the system. By implication, Systems Biology focuses on the non-linear interactions between all the components that contribute to a function of interest [1]. The multitude of components in biological systems and the complexity of their interactions requires mathematical modeling in the context of quantitative experimental assessment of the interactive properties of the components and of system functioning [3].

Probably due to natural selection in favor of complex functions and redundancy to provide robustness, biochemical networks appear to be much more complex than their usual chemical counterparts. This includes the complexity of the interactive properties of their components, e.g. allosteric, activatory and inhibitory interactions. On the other hand, thanks to the dominance of a single type of catalyst, i.e. the protein, encoded by a single type of molecular template, i.e. a gene, the diverse properties of the components

of biochemical networks are perhaps more traceable to material properties that can be determined for the components in isolation, than are components of chemical networks. It is likely therefore that mechanistic modeling relating systems properties to components properties will be especially rewarding for biochemical networks.

2.2 Chemical reaction networks

The work by Heinrich and colleagues led to a change in the approach to the modeling of metabolic pathways. Before that time, metabolic networks were modeled as a large set of chemical processes, typically in a continuously stirred tank reactor, where a variety of substances are created and degraded. This standard approach employed ordinary differential equations such as:

$$\frac{\partial c_i}{\partial t} = f_i(c_i, c_j, p_m) \tag{1}$$

where the rates of change of any substance (i) was considered an explicit function of the concentration of itself (c_i) (typically to first order), of the other substances (c_j), and of a large number of parameters (p_m) including kinetic constants and external conditions such as temperature and pressure. It may be noted that we here use the 'curved d' ∂ for the time derivative. Indeed, we look upon the concentration as if its is a function of all parameters in the system, including a parameter that is called time. Below the utility of this perspective will become clear as we shall deal with control by time in the same vein as control by other parameters.

If parameters were known, or could be estimated or hypothesized, the differential equations were integrated numerically, leading to:

$$c_i(t) = F(c_i(0), c_j(0), p_m, t) \tag{2}$$

if the solution was unique. The predicted time dependence for the substances could then be compared to experimentally observed behavior, and parameter values could be adjusted so as to improve the fit between prediction and observation. The complexity of this approach was limited by reducing the complexity of the differential equations and of the number of components.

The idea was that once the model was performing satisfactorily, the sensitivity of the solutions with respect to the parameters might be determined, enabling the engineer to change temperature or pressure or perhaps even to add a catalyst to enhance one of the rate constants. Sensitivity coefficients might be defined as:

$$S_{p_m}^{c_i,t} = \frac{\partial \ln c_i(t)}{\partial \ln p_m} \tag{3}$$

with infinite times for the steady state versions.

There would be no particular urge to inspect the values of the sum of these sensitivity coefficients (for all parameters), but if one were to do so, no specific number (such as 1 or 0) would come out, not even at steady state. There might be some motivation to look

at the sensitivity coefficients with respect to temperature or pressure, as these parameters values can often be altered by the experimenter or the engineer, such that predicted improvements of the process can indeed be tested or put to practice. One may also be interested in calculating the response to changes in the kinetic parameter values, as this reflects the importance of the corresponding processes for the phenomenon of interest.

2.3 The special invariant properties of biochemical reaction networks

In most biochemical pathways, the situation is different. These consist of a limited set of specific reactions, each catalyzed by a specific enzyme, the concentration of which can be changed in principle, and the kinetic properties of which can often be determined in isolation. Each substance is produced and consumed by a limited number of these reactions at well-defined stoicheiometries. The change in concentration of each component can then be written in terms of a number of reaction rates (υ) and stoicheiometric coefficients:

$$\frac{\partial c_i}{\partial t} = \sum_{j=1}^{n} N_{ij} \cdot \upsilon_j(k_{\mathrm{cat}\,j}, c_k, e_l, K_{Mj}, P, T) \tag{4}$$

The N's represent the stoicheiometric coefficients, positive when referring to the product, and negative when referring to the substrate of the reaction. P, and T represent physical parameters such as Temperature and Pressure, e_l represents enzyme concentration, $k_{\mathrm{cat}\,j}$ represents the catalytic rate constant of the enzymes whereas K_{Mj} represents all remaining kinetic parameters that occur in the enzyme rate laws.

When addressing non-channeling metabolic pathways, the biochemical paradigm goes even further than the above equation. Enzymes tend to be quite specific. Hence a reaction rate in a living cell tends to depend only on the concentration of the single enzyme that catalyzes them, not directly on any of the other enzymes (except in case of isoenzymes). Moreover, the enzyme is a catalyst implying that the rate equation can be written as the product of a function that depends on the concentration of the enzyme and a rate function that becomes zero at thermodynamic equilibrium and is independent of enzyme concentrations. In many cases the rate is proportional to enzyme concentration [9]. If it is not proportional, then it pays off for simplicity to define an enzyme activity e for the reaction such that the rate is proportional to the concentration of that, then virtual enzyme. The above equation then reduces to:

$$\frac{\partial c_i}{\partial t} = \sum_{j=1}^{n} N_{ij} \cdot e_j \cdot \upsilon_j(c_k, K_{Mj}, P, T) \tag{5}$$

Many modelers read this equation again as an ordinary differential equation, amalgamate the parameters so as to effectively simplify it, and thus return to Eq. 1. They may then integrate or otherwise determine its properties (e.g. stability by inspecting the eigenvalues of the Jacobian) as a function of the amalgamated parameters. This may then again determine the sensitivity of the behavior with respect to some of the amalgamated parameters. The approach may then serve to show what type of dynamic

behavior (e.g. oscillations versus steady state) the system could exhibit for all possible values of the amalgamated parameters.

Although technically their computers integrated the differential equations, biochemists led by the school of Heinrich and Rapoport did not just treat the system in terms of differential equations with amalgamated parameters. They rather kept track of the rate equations and of their parameters with biochemical meaning, such as the enzyme concentrations. At least some biochemical reaction steps in their models were and remained explicit. For Systems Biology, which tries to relate systems behavior to the properties of components and their interactions, this is highly important.

The modeling of metabolic pathways by modeling of enzymatic steps explicitly, was not invented in Berlin. Groups such as those around Britton Chance, including Garfinkel, Kohn, Higgins and Wilson, had engaged in this before. The rate equations they used did refer to the steps in the pathways. In the case of Garfinkel this led to ambitious models of very large chunks of metabolism [14]. However because information was lacking about the kinetic parameters and rate laws of the biochemical reactions, mass action laws were often used and parameters values were obtained by fitting to experimental behavior of the system rather than of the individual reaction steps. The models were underdetermined, i.e. they had too many parameters as compared to the limited richness of the experimental dataset. In some cases a number of parameter values were fixed provisionally on the basis of 'reasonable estimates'. In other cases parameters were again amalgamated. A further step in this direction was the use of phenomenological or amalgamated rate equations, such as when using a Michaelis–Menten equation for the dependence of mitochondrial respiration on the concentration of ADP [6]. In the end therefore this modeling activity slowed down, essentially because of lack of the appropriate experimental information.

3 The biochemical reactions in networks

The mass of a living cell equals the sum of the masses of its components. The component masses are essentially independent of whether the components are inside the cell or in a test tube. Mass is a linear property, where the word 'linear' refers to these features of additivity and autonomy. The rate at which cells consume glucose is also the sum of all the individual rates at which the cellular processes consume glucose. However, the latter rates would not be the same if the individual glucose consuming reaction were to be carried out in a test-tube: most of the intracellular processes influence each other; reaction rates are therewith 'non-linear' properties'. Equation 4 also shows that the behavior in time of the concentrations of all the system components is a function of the rates of all the processes, which again change as those concentrations change in time. Accordingly, the process rates are important dynamic components of cell biology. Cell biology may be seen as a function of the rates of all the processes in the cell, much more than a sum of all the masses, or crystal structures, of the components. It makes sense to look at the processes in terms of being the most important components of the cell. Equation 5 pinpoints that these components often relate to material components in the sense of enzymes. What is important then is to understand the concentrations of these enzymes and the equations for the rate at which they carry out their process. We shall now discuss these 'rate equations'.

3.1 Precise rate equations

Two types of enzyme rate equations may be distinguished, i.e. the mechanistic type and the phenomenological type. In the former type, the equations still reflect the inner workings of the enzyme that is being modeled. For an enzyme this means that a rate equation is used that is based on the known action mechanism of the enzyme in the pathway. Parameters in the equation are mechanistic at that level of description, i.e. their meaning is understood in such terms. If the enzyme reversibly binds a single molecule of substrate (S) and then in a second reaction step converts this irreversibly to a single product, then at the quasi-steady state for the enzyme, the mechanistic rate equation would contain the three corresponding kinetic constants (k_1, k_2 and k_{-1}):

$$v = \frac{k_1 \cdot [S] \cdot k_2 \cdot e}{k_1 \cdot [S] + k_{-1} + k_2} \tag{6}$$

The rate constant of the reaction of binding of substrate by enzyme is represented by k_1, the rate constant of the reaction of dissociation of enzyme-substrate complex is represented by k_{-1} and the rate constant of the reaction of formation of product is represented by k_2. (If the enzyme kinetic mechanism were to be described in complete detail then this equation would be more complex; k_2 is a composite of a large number of more elementary transition probabilities between enzyme states.) The corresponding, precise yet phenomenological rate equation is the Michaelis–Menten equation:

$$v = \frac{[S] \cdot V_{\max}}{[S] + K_M} \tag{7}$$

$V_{\max}$ is defined as the rate at infinitely high concentration of the substrate and K_M is numerically equal to the substrate concentration at which the actual rate becomes half this maximum rate.

Rate equations for the individual steps in a metabolic network can serve multiple functions in the modeling and understanding of those pathways. First, an important function is to understand how that step itself contributes to and influences the functioning of the pathway. This is in turn fully determined by the ways in which the enzyme 'talks' to the pathway in language that could be heard by the latter, plus all the ways in which the enzyme 'listens' to the pathway and its environment. The 'listening' is the adjustment of its talking to messages it receives from the pathway. The enzyme talks to the pathway by producing molecules and consuming molecules at a certain rate. This talking is characterized by the rate (v_j) of the reaction the enzyme is catalyzing and the stoicheiometric coefficient (N_{ij}) for each metabolite. For, consuming their substrate and producting their product is the only acitivity of the many the enzyme may carry out (such as changing its comformation, binding an allosteric modifier, being phosphorylated by a kinase) that affects the rest of the system and the other molecules within it. The listening consists of the extent to which the enzyme adjusts the rate in response to the changes in its environment that the enzyme can actually perceive. For small changes in the enzyme environment the listening is indicated by the 'so-called' elasticity coefficients of the enzyme with respect to all the metabolites

that affect the enzyme and by corresponding additional elasticity coefficients for non-metabolite influences, such as Temperature. The elasticity coefficients are defined as the partial log-log derivatives of rate with respect to any of the arguments of the rate function. For instance, the process we described in Eqs. 6 and 7 would have elasticity for its substrate S of:

$$\varepsilon_S^v \equiv \frac{\partial \ln v}{\partial \ln[S]} = \frac{K_M}{[S] + K_M} = \frac{1}{1 + \frac{k_1 \cdot [S]}{k_{-1} + k_2}} \tag{8}$$

For this particular rate Eq. 6, the enzyme listens carefully (elasticity coefficient approaching 1) when the substrate concentration is well below the Michaelis constant, but not so well when it is far above. For larger changes in the environment of the enzyme, the listening is characterized by its rate equation and the parameters in that equation.

As shown by Eq. 8, both the mechanistic and the phenomenological enzyme kinetic equation can serve to understand the listening function of the enzyme in the pathway. In fact the phenomenological rate equation has an advantage here as it requires the knowledge of fewer parameters, which are also more readily accessible experimentally. In our example only $V_{\max}$ and K_M are required to represent the talking and the listening, both of which can be determined by measuring the overall enzyme reaction rate as a function of the concentration of the substrate. The determination of k_1, k_2, and k_{-2} would require fast kinetic experiments plus determination of the enzyme concentration. In fact, although of interest for understanding the catalytic mechanism of the enzyme, the magnitude of these parameters is not of interest for understanding the regulation of pathway functioning. This reflects the more general feature that provided that the parameters of the phenomenological rate equation are (and need to be, as a priori calculations are generally too inaccurate) determined experimentally, most of the detailed properties of enzymes, including their 3-D structure, is irrelevant for understanding how they determine the functioning of the pathway.

A second important function of the enzyme kinetic rate equation is to enable one to understand the molecular mechanisms through which the environment of the pathway influences the functioning of the pathway through the components of the pathway. An important case here is the influence that alterations of gene expression may have on pathway function through changes in the concentrations of pathway enzymes. The phenomenological, Michaelis–Menten equation (Eq. 7) by itself is not suitable for this. Because it contains the enzyme concentration explicitly, the mechanistic equation (Eq. 6) does enable one to begin to assess the effect of this external influence. In practice this limitation is circumvented by utilizing the relationship that $V_{\max}$ is proportional to enzyme concentration (if it is) in addition to the phenomenological equation, but this of course means that in practice one seeks a best possible compromise between the advantages and disadvantages of using phenomenological versus mechanistic kinetic equations.

Both the mechanistic and the precise phenomenological rate equation may be rephrased, i.e. described as functions of properties that are related to the concentrations of substrates and products. The concentrations of free substrate and of free product that occur in the equations may be replaced by the exponential function of

their thermodynamic potential [41]. We here discuss the case of a linear metabolic pathway in which the second step is an enzyme catalyzed conversion of metabolite X to a molecule of metabolite Y. In order to keep this illustration simple, we shall assume that the reverse maximum rate is small enough to be neglected, but we retain the product inhibition term. The precise, phenomenological, rate equation for this reaction is:

$$v = \frac{\frac{[X]}{K_X} \cdot V}{1 + \frac{[X]}{K_X} + \frac{[Y]}{K_Y}} \tag{9}$$

$[X]$ and $[Y]$ refer to the activities (free concentrations) of the two corresponding metabolites. The chemical potential for X (μ_X), its standard chemical potential ($\mu_X^{0\prime}$), and its activity $[X]$ are related through:

$$[X] = e^{\ln[X]} = e^{\left(\mu_X - \mu_X^{0\prime}\right)/RT} \tag{10}$$

One may write the rate as a function of the two chemical potentials:

$$\frac{v}{V} = \frac{\frac{e^{\mu_X/RT}}{\kappa_X}}{1 + \frac{e^{\mu_X/RT}}{\kappa_X} + \frac{e^{\mu_Y/RT}}{\kappa_Y}} \tag{11}$$

Where:

$$\kappa_X = K_X \cdot e^{\mu_X^{0\prime}/RT} \tag{12}$$

and

$$\kappa_Y = K_Y \cdot e^{\mu_Y^{0\prime}/RT} \tag{13}$$

This equation corresponds to the precise non-equilibrium thermodynamic description of the reaction, showing how the two thermodynamic driving forces μ_X and μ_Y push and inhibit, respectively, the flow through the enzyme [41].

3.2 Approximate kinetic equations

Equation 5 can be integrated numerically. Therewith the behavior of the system can be predicted precisely (cf. below) provided the functional form of the rate equations and the parameter values are known. Only for very few pathways (http://www.siliconcel.net; http://www.jjj.bio.vu.nl/) this is the case. Most often there is considerable uncertainty in parameters and this necessitates sampling of the parameter space in order to assess the range of model outputs or for inverse modeling strategies (see below). These search spaces grow combinatorially with the number of parameters. In this

context, especially for larger intracellular networks, it is worthwhile to try to approximate the exact rate equations with approximate equations that have fewer parameters or variables. Such reduced models can be integrated more rapidly and, even more importantly, have much smaller search spaces that may be more fully explored.

Another valid motivation for the development of approximate kinetic schemes is to provide analytic solutions that give insight into the critical network interactions or that relate more directly to a scientific issue that is being examined. Finally, there has historically been a tendency to develop unified kinetic frameworks that use a specific functional form for the rate equations. This approach provides some degree of mathematical expediency for general analysis and can be valuable for systems in which the actual rate expressions are still unknown. We shall here discuss approximate equations that fall into some of the categories discussed above.

A general approach is to approximate the rate equations by linear expressions in terms of, possibly non-linear functions (g_k) of the concentrations:

$$v_j = \sum_{k=1}^{n} m_{jk} \cdot g_k(c_k, e_j, K_{Mj}, P, T) + a_j \tag{14}$$

m_{jk} are proportionality constants not identical to the stoicheiometric coefficients. The most frequently used approximation of this type is the linearized mass action approximation, where the function g_k is simply the concentration itself. This approximation makes the solution of the steady-state equation for the overall system particularly simple and also allows the analytical solution of the time dependence of the system. However, enzyme reactions tend to saturate at higher concentration of the substrate, and they are often allosterically regulated and both these features have considerable impact on the behavior of the network. Linearized rate equations have therefore not proven to be very useful for the description of the actual system properties of biochemical networks. Although almost equally amenable for integration for systems at steady state, non-linearized mass action equations of this type may well be quite useful.

Another possibly useful simplified rate equation of the above type may be:

$$v_k(X, Y) = \frac{[X]}{[X] + K_X} \cdot V_k - \frac{[Y]}{[Y] + K_Y} \cdot V_{-k} \tag{15}$$

which is similar (but not identical) to an irreversible product insensitive Michaelis–Menten equation. X refers to the substrate of the reaction, and Y to its product. It should be noted that the same K_X should accompany X in all rate equations where X occurs, which is not often realistic. Below we shall show that this approach offers some potential.

Another approximation writes the rate as a linear function of the chemical potentials of the metabolites, which corresponds to the logarithm of their concentrations. This approach has been called (mosaic) linear non-equilibrium thermodynamics [42], or lin–log kinetics [48]. An early approach to the use of approximate, phenomenological rate equations was classical non-equilibrium thermodynamics, which postulated linearly proportional relations between rates and Gibbs energy differences cross reactions. The basis for the postulated linearity was not more than neglecting the higher

order terms of the corresponding Taylor expansion of the rate equation plus Onsager reciprocity, which is valid only close to equilibrium. When the reactions are away from equilibrium, as is the case for many important reactions in living cells, then the Onsager reciprocity tends to be absent but linear approximations of the relations between rates and the individual chemical potentials of the reactants and products were descriptive of experimental observations [39]. The corresponding slopes were called 'absolute elasticity coefficients (see [42] for an overview).

Linearizing the above, precise non-equilibrium thermodynamic rate equations (Eq. 11) for enzyme kinetics around some state, which will be referred to by an asterisk, one obtains the relationship:

$$\begin{aligned} v &= v^* + \gamma_X \cdot (\mu_X - {\mu_X}^*) - \gamma_Y \cdot (\mu_Y - {\mu_Y}^*) \\ &= v^{\text{£}} + \gamma_X \cdot R \cdot T \cdot \ln[X] - \gamma_Y \cdot R \cdot T \cdot \ln[Y] \end{aligned} \tag{16}$$

Here:

$$\gamma_X = V \cdot \frac{\frac{[X^*]}{K_X} \cdot \left(1 + \frac{[Y^*]}{K_Y}\right)}{\left(1 + \frac{[X^*]}{K_X} + \frac{[Y^*]}{K_Y}\right)^2} \tag{17}$$

$$\gamma_Y = V \cdot \frac{\frac{[X^*]}{K_X} \cdot \frac{[Y^*]}{K_Y}}{\left(1 + \frac{[X^*]}{K_X} + \frac{[Y^*]}{K_Y}\right)^2} \tag{18}$$

Equation 16 is the linear non-equilibrium thermodynamical rate equation for the reaction. It may be rewritten as a relationship between the rate and the Gibbs energy difference across the reaction plus a term that depends on the chemical potential of X as extra driving force:

$$v = v^{*\prime} + \gamma_Y \cdot (\mu_X - \mu_Y - ({\mu_X}^* - {\mu_Y}^*)) + \gamma_{XX} \cdot \left(\mu_X - {\mu_X}^*\right) \tag{19}$$

This relationship can be useful because it relates the kinetic behavior of the enzyme and system to the thermodynamics and may help prevent violations of the second law of thermodynamics that might occur otherwise.

Because the chemical potential is a linear function of the logarithm of the activity or concentration of a substance, the above equation for v also corresponds to the so-called lin–log kinetic equation for this case. In some cases, e.g. when the sum of $[X]$ and $[Y]$ is constant, the second derivative of the rate with respect to the chemical potential equals zero at a certain concentration of the two, giving rise to a more than trivial range of linearity [39,42], In general however, the linearity described by Eq. 17 is nothing but approximating the true rate equation by its tangent.

Not all power-law approximations follow the paradigm of Eq. 14. Biochemical Systems Theory has introduced power-law approximations for the rate equations [37]:

$$v_k(X, Y) = \alpha_k \cdot [X]^{\beta_{kx}} \cdot [Y]^{\beta_{ky}} \tag{20}$$

If Y corresponds to a product of the reaction, then its inhibitory effect is reflected by its exponent being negative.

4 Inverse and forward modeling

Integration of Eq. 5 gives the concentrations of the system components as a function of time. For the 'metabolic systems' [42] that are the paradigm for living cells (i.e. systems that are open in the thermodynamic sense for some but not all metabolites) the state variables are the independent metabolite concentrations. Often only steady states are considered, and the state variables at steady state are determined by the rates balancing such that the corresponding time derivatives become equal to zero (i.e. valid for systems at mechanical and thermal equilibrium with their environment: only mass conservation matters):

$$0 = \sum_{j=1}^{n} N_{ij} \cdot e_j \cdot \upsilon_j(c_k, K_{Mj}, P, T) \tag{21}$$

So-called 'forward modeling' starts from the known, or supposedly known parameter values and known, precise or approximate rate equations for the components of the network. It then integrates these, through Eq. 5 or 21 to obtain the time-dependent or steady-state properties of the system, respectively. The latter may then be compared to the system behavior that is observed experimentally and if the two do not correspond, the model (with its parameter values) is considered falsified.

One may then try to see if for different parameter values the predictions of the model do correspond to the observations, i.e. one inserts a new set of parameter values into the model, reintegrates and asserts whether the fit to the experimental results is better.

Guessing the better parameter values is difficult however. The problem of finding the models that predict the observed functional behavior is called inverse modeling. Here we shall focus the discussion on Forward Modeling, with minor reference to inverse modeling at the end.

4.1 Forward modeling: integrating the interactive properties of molecular components to understand systems behavior

In principle, the dynamic behavior of the variable metabolite concentrations in the network should be predicted by the integration of Eq. 5 (or the solution of Eq. 21). The integral is a function of all enzyme activities and of all other parameters in the equation, inclusive of those that measure the interactions of the enzymes with the metabolites and hence through the metabolites with each other. Therewith this integration brings about one of the aims of systems biology, i.e. to relate functional behavior of the system to its molecular and interactive properties. We shall here discuss three different approaches [1]:

1. Integration of the precise rate equations,
2. Integrating the system equations after simplifying the enzyme rate equations,

3. Simplifying and limiting the questions asked; alternative conceptual ways of describing system behavior.

4.1.1 Integration of the precise rate equations; silicon cell and JWS online

Because the differential equations are non-linear, they cannot usually be integrated analytically. The silicon cell-JWS Online approach [30,38], inspired by the erythrocyte glycolysis model of the Berlin group (for a recent version see [22], which is available on JWS; JWS stands for JJava Web Simulation), accepts this complexity and integrates the differential equations numerically. Fluxes and concentrations can be calculated as a function of time, also after changing any of the parameters values. Comparison with experimental results should constitute hard tests of the correctness of the model underlying the silicon pathway/cell and of the experimentally determined parameter values. And, in the case of JWS-silicon cell, this can be done by anyone in the world, through a world wide web interface (http://www.jjj.bio.vu.nl/). The in silico model may help understanding, as in silico it is much easier to modulate the system under study and calculate what the implications of the modulations are, and, one may even calculate the extent to which regulation runs through the various links in the network.

Of course there are limitations. For only very few pathways the values of the parameters characterizing the molecular interactions are known at sufficient accuracy. Where knowledge exists of kinetic parameters such as the Michaelis constants and k_{cat}'s, V_{max} values are often uncertain because expression levels are not known quantitatively enough. In addition in some cases allosteric effectors may be active in vivo, or enzymes may be modified covalently, which may not have been taken into account in the in vitro assays, causing a difference between the situation in vitro and that in vivo. And then, the result of the numerical integration by itself is less rewarding intellectually than it may seem; it may lead to predictability, but not to something perceived as 'understanding' by the human mind. Understanding may require simplified models that show essence rather than detail; even though the use of such models alone is not satisfactory. Such models may show what mechanisms could be responsible for an important function of the system, rather than what mechanism actually is responsible.

4.1.2 Integrating the system equations after simplifying the enzyme rate equations

Simplifications of the enzyme kinetic rate equations becomes particularly rewarding if they enable the analytical solutions of the equations that express the behavior of the pathway, as an analytical function of properties of the enzymes.

The simplified rate equations of the type of Eq. 14, enable analytical solution for the steady state, as Eq. 5 becomes:

$$0 = \sum_{j=1}^{n} N_{ij} \cdot e_j \cdot \left(\sum_{k=1}^{m} m_{jk} \cdot g_k(c_k, K_{Mj}, P, T) + a_j \right) \tag{22}$$

which is linear in g_k. Indeed, when modeling the flux through, or concentrations in the pathway there is an appreciable computational advantage in using the linear

non-equilibrium thermodynamics (lin–log) approximations: the equations are linear in the chemical potentials and steady-state solutions can be obtained algebraically (see above and [26,42,48]. These solutions will describe the operating state (by definition) and minor variations around that state, where the extent of variation is limited by the accuracy of the lin–log approximation. The error made by the lin–log approximation may often be insignificant as compared to the magnitude of the experimental error. The use of the precise rate equations has the disadvantage that most often numerical solutions are required and this is why non-equilibrium thermodynamics and lin–log approximations have been favored occasionally.

The power-law approximations also lead to an elegant way of solving for the steady state. Each time change in concentration (the left-hand side of Eq. 4), is then written as the difference between an afferent term and an efferent term. Both of these then correspond to a power law, provided that they consist of single reactions. If there is more than one afferent reaction or more than one efferent reaction, the sum of the former or the sum of the latter need to be each approximated by a new power law. Then the steady state requires that two power laws become equal. Taking the logarithm of these equations, one then finds a linear equation that relates the logarithms of the concentration of all the components of the system. The solutions then become quite similar to those of the non-equilibrium thermodynamic/lin–log approach [37].

In much of mathematical biology complicated rate equations are approximated by simpler equations. There are two essentially different bases for such simpler equations. One is phenomenological: an equation which is known to be more helpful for the subsequent systems analysis, is taken and fitted to the dynamic behavior of a component. The above mentioned power-law and lin–log approximations belong to this category. The other category of equations is mechanistically precise, the simplification (or rather) being based on reduced dimensionality of the behavior under the condition relevant for the analysis. An example is the description of the kinetic behavior of an enzyme, which could be done stochastically taking into account the complete molecular dynamics of the movement of all its atoms. There is no doubt that the molecular dynamics leads to fluctuations in binding, dissociation and conversion rates of the substrate and product of the enzyme catalyzed reaction. Yet at the time scale that is often more relevant for fluxes through metabolic networks, these fluctuations are often averaged out. Descriptions in terms of average rate constants for only a few transitions of the enzyme between average states are then both accurate and based on the (averaged) actual molecular behavior. Another example is the irreversible, product inhibited Michaelis–Menten equation that may be used for conditions where observation is much slower than 1 second, the Gibbs energy drop across the reaction exceeds 12 kJ/mol, and the total enzyme concentration is more than 100 times lower than the concentrations of its substrates and products. In this case the constants such as K_M and $V_{\max}$ are identifiable functions of the mechanistic rate constants, hence fixed. They are identified by assuming quasi-steady state for the enzyme thus removing degrees of freedom (reducing dimensionality of described behavior).

Both approximations may come within the same error margin of less than 20%. Yet they then have greatly different qualifications. The form of the simpler mechanistic equations is only accidentally optimal for systems analysis. More often the phenomenological equations offer the advantage that they have been selected to enable simpler,

often even analytical solutions of that behavior (cf. above), or they may otherwise simplify the analysis. Possible types of systems-level mechanism that could underlie complex biological phenomena can be spotted in this manner. Examples may be found in the analysis of glycolytic oscillations where various, simplified models have indeed helped to pinpoint the possible role of product stimulated phosphofructokinase and of feedback loops in the glycolytic pathway [15]. Another example has been important in developmental biology where with simplified models it was shown how pattern formation in early development could readily be explained by a simple set of non-linear interactions between cells [29]. However, in the case of glycolytic oscillations two different types of pathway organization may lead to the oscillations, and it was shown that most likely the oscillatory tendency was controlled by many components in the pathway at the same time [36]. In the case of pattern formation in Drosophila, the proposed mechanism of self-organization was not responsible for the formation of the stripes [28]. It is one thing to show that a type of mechanism could be responsible for an important biological phenomenon. It is something else, and more important to show that that mechanism is actually responsible. To achieve the former, approximate, simpler, phenomenological rate equations may well suffice. To achieve the latter, the analysis must probably be more closely related to actual mechanism than the phenomenological equations permit. Then simplification should perhaps be limited to mechanistic simplifications.

Indeed there are a number of major drawbacks in the use of phenomenological equations for the kinetics of the component enzymes, as compared to the mechanistically precise rate equation approach. First, when the system remains the same in terms of its composition, hence in terms of the concentrations and kinetic properties of its enzymes, but functions around a significantly different concentration of pathway substrate or product, then the approximation will tend to fall outside its range of applicability. One should then fit the parameters to the kinetics of the enzyme at this new operating state, but this usually requires much extra and tedious experimentation. If one does so, then all or most parameters will have changed, in the absence of any changes in the molecular properties of the macromolecules in the pathway. This is a bit unsatisfactory as the resulting model then really functions as a set of hardly related ad hoc descriptions, one for each operating state. When the mechanistically precise equations are used, the parameters should remain the same; the description should remain valid.

Second, because one usually does not re-determine the new approximate rate equation for the enzyme in its new operating state, but rather re-fits the parameters in the approximate kinetic description to make the system predict the system behavior in both operating states, the system model tends to become unstable: As the data set to which the model is fitted, changes, the best fit and hence the model parameters change in terms of magnitude, even though the molecules in the system do not change. This is still an aspect of many models of cell biology such as models of the cell cycle [7]. And it becomes an ever increasing problem as the model and the total experimental dataset gets larger, making it more and more difficult to find the best fit. An important aspect here is that the rate equations and parameter values have no basis other than that they fit system behavior. This implies that there is no reason why they should be the same under two different conditions of the system, even if the properties of the macromolecules in the system are the same.

Third, the approximate phenomenological approach tends to be non-predictive and non-falsifiable. In the approach with the mechanistically precise equations, the effect of changing a Michaelis constant through a change in isoform of any of the enzymes can be predicted by the model. When the prediction fails, the model used is falsified and a new network diagram may have to be drawn for the system. This is less straightforward in lin–log kinetics, although it could be done by adjusting its parameters by using the relations between them and the Michaelis constants (e.g. above); however in this case all parameters of the lin–log relationship would change as opposed to a single change in the mechanistically precise rate equation. Consequently, the statistical basis for any falsification or verification would be compromised.

Fourth, the approximate approach tends to be non-inductive. Because the parameters of the lin–log equations are often fitted to make the equations describe the behavior of the pathway, fitting parameter values may well be found that do not correspond to true properties of the macromolecules of the system or their interactions. Unknown but active interactions may not be discovered. The mechanistically precise equation approach tests whether the pathway behavior can be understood on the basis of the actual properties of the macromolecules. If there are macromolecules or interactions active that are not represented in the kinetic model, the precise equation approach should fail to model the behavior of the pathway.

Modeling biochemical reaction networks can have many different purposes. One of these may be to interpolate between data points. Another purpose may be the effective storage of information without the noise. The one we are here focusing on has the aim of understanding the functioning of the network in terms of its molecular components and their interactions, or conversely, of learning more about those components of the network from observed network behavior. The former aim may include the ambition to predict the effect of a mutation or a drug on system function, the latter of suggesting a molecular mechanism for a disease.

The mechanistically precise-rate-equation ('Silicon cell') approach is not practiced much, largely because the required experimentation is often tedious if not impossible and because the number of alternative ways to explain the fallacy of the model is too large for all of them to be chased. In such cases the more phenomenological approach is the one to implement: Both approaches have their merits. Good modeling practice requires one to specify which of the two approaches is being followed. In many existing models, the documentation of how the model was evolved, is still incomplete.

Also the functionality of the two types of model is different. If experimental behavior of the system does not correspond to the predictions by a phenomenological model, one may adjust (fit) parameters until it does correspond, and then conclude that the phenomenological model works. By contrast, when using a mechanistically precise model, one should first conclude that the model (or the experiment) is wrong, and try to figure out which molecular component has escaped attention, which interaction has been overlooked, or which interaction that is incorporated in the model may be wrong. In this procedure, one may provisionally examine which parameter fits would generate correspondence between model and experiment *and then check by experimentation* with the relevant components whether the altered parameter value or property can be confirmed with the relevant components in vitro. An example of examining whether additional properties make the model fit the experimental results

better, could be that of enabling two enzymes in the model to dimerize and reduce each others V_{max}. If the fit would be better, then this dimerization should be established with the two enzymes in vitro and at the relevant concentrations, before it is entered into the precise model. Another example could be the discovery that assuming the presence of a shunt reaction in the network would improve model behavior. This example is real: Rapoport and coworkers did discover the importance of the PGA shunt in erythrocytes [34].

4.1.3 Simplifying and limiting the questions asked

The modeling discussed up to this point was directed at understanding or describing the intracellular network of interest completely. With 'completely' we here refer to complete knowledge of all state functions as a function of time, hence complete understanding of the system. Usually one is not interested in all aspects of a system. Already above, we discussed approaches that were less complete, i.e. limited to systems at steady state. We shall now discuss further limitations to the questions asked about the systems. One is the case where one agrees only to be interested in the fluxes and not in the concentrations of the substances. Another one is where one is interested only in the control of the system.

4.1.3.1 Flux analysis and flux balance analysis One of the simplest approaches taken is to interpret Eq. 21 as linear equations in the steady state rates v and to only be interested in the rates that satisfy the above equation, i.e. the steady state conditions. Some of the rates correspond to fluxes into or out of the system and can be determined by measuring changes in external metabolite concentrations. Internal rates at steady state can then be solved from Eq. 21, except that there are usually more rates v than independent metabolite concentrations, i.e. the above equations are fewer in number than rates v. For a linear pathway, with m metabolites, there are m equations but $m + 1$ rates. However, the steady state flux through the pathway is often known or can be measured by measuring the disappearance of external substrate or the appearance of external product. Hence for steady states the problem can be solved, in an approach called flux analysis.

In the case of large networks, this may not suffice because they have parallel pathways. Here an approach called flux balance analysis selects the pathways that cost the least ATP or generate most ATP, requiring maximum energetic yield [33]. Even then there is often more than a single solution for all the rates, but this may not matter, as these solutions may be functionally equivalent in terms of ATP production although not in other functions. A problem with flux-balance analysis is that it is quite uncertain whether the assumption of optimum efficiency is realistic. In fact it is known for simple organisms like yeast that they often do not choose the pathway that is most efficient thermodynamically [42].

4.1.3.2 Metabolic Control Analysis for Systems Biology: Fundamental principles of the control of intracellular networks Much of Systems Biology began with genome-wide approaches accumulating large amounts of data concerning macromolecules in living organisms. The aim was to observe patterns in the dynamics of the concentrations

of these macromolecules and to deduce from these patterns how these macromolecules are related, e.g. by belonging to the same pathway. This Systems Biology is rooted in Genomics and Molecular Biology. A second root of Systems Biology lies in mathematical biology and physical chemical approaches such as kinetics and non-equilibrium thermodynamics. The corresponding Systems Biology examines how and why systems of macromolecules function differently from how the individual components would act in isolation [44]. It is here that Mathematical Biology plays a crucial role [41].

Metabolic Control Analysis and its expansion Hierarchical Control Analysis are amongst the approaches that lead to understanding beyond precise prediction of behavior. This may be illustrated by querying to what extent any of the enzymes in a linear metabolic pathway of 14 enzymes determines their own flux. In isolation the rates of the enzyme catalyzed reaction is proportional to the amount of enzyme, hence the control coefficient of the enzyme is 1 and it is 100% in control of its own flux. Would that be the same when the enzyme sits in the pathway? Is there a general principle that gives some insight in this? Is there a methodology to understand the extent of control?

Metabolic Control Analysis (MCA) provides the answers to these queries as well as the required methodologies. For this it defined the flux control coefficient. We shall use a rather systematic way of defining the flux control coefficients. In order to do this, we re-emphasize that cell behavior can be seen as a simultaneous function of all processes. We introduce a modulation parameter for each individual process, as a multiplier in its rate equations. We also introduce a modulation parameter for time. For the balance equations this leads to:

$$\frac{\partial c_i}{\partial\,(\lambda_t \cdot t)} = \sum_{j=1}^{n} N_{ij} \cdot \lambda_j \cdot \upsilon_j(c_k, e_l, k_{\mathrm{cat},l}, K_{Mj}, P, T) \tag{23}$$

Whenever Eq. 5 applies:

$$\lambda_j = e_j \tag{24}$$

Equation 23 shows that in unique asymptotically stable systems a state function f that depends on the concentrations of the substances in the system, is a function of all process activities, and of some time-independent external and internal parameters and of the initial conditions:

$$f = f(t, \lambda_t, \lambda_j, e_l, k_{\mathrm{cat},l}, K_{Mj}, P, T, K_M, c_i(0)) \tag{25}$$

The control exercised by any process j on any system function f is then defined as the corresponding control coefficient:

$$C_j^f \equiv \frac{\partial \ln f}{\partial \ln \lambda_j} \tag{26}$$

The control by time is defined likewise:

$$C_t^f \equiv \frac{\partial \ln f}{\partial \ln t} \tag{27}$$

The dependences with respect to the other arguments of the function f are defined as Response coefficients, e.g.:

$$R_{K_M}^f \equiv \frac{\partial \ln f}{\partial \ln K_M} \tag{28}$$

For f we shall first consider any of the independently varying metabolite concentrations themselves. We look at two systems, i.e. one with all lambda's equal to 1 and the concentration equal to c, and a transformed system where the concentrations are given by c' and where from time zero onward:

$$\lambda_i = {}^1\!/_{\lambda_t} = \lambda \tag{29}$$

Inspection of Eq. 23 shows that:

$$\frac{\partial c_i}{\partial t} \equiv \frac{\partial c_i'}{\partial t} \tag{30}$$

As at time zero the concentrations in the standard and the transformed system were equal, also:

$$c_i(t) \equiv c_i(t)' \tag{31}$$

Since this is true for all extents of modulation:

$$0 = \frac{\partial \ln(c_i)}{\partial \ln \lambda} = \frac{\partial \ln(c_i)}{\partial \ln \lambda_t} \cdot \frac{d \ln \lambda_t}{d \ln \lambda} + \sum_{j=1}^{n} \frac{\partial \ln(c_i)}{\partial \ln \lambda_j} \cdot \frac{d \ln \lambda_j}{d \ln \lambda} = -C_t^{c_i} + \sum_{j=1}^{n} C_j^{c_i} \tag{32}$$

This is the time dependent summation law for concentration control coefficients [23]. It appears to be a fundamental aspect of all dynamic systems. At steady state the time dependence disappears and one obtains the concentration-control summation law of Metabolic Control Analysis [19,24].

For any rate one finds:

$$v_i(t)' = \lambda_j \cdot \upsilon_j(c_k, e_l, k_{\text{cat},l}, K_{Mj}, P, T) = \lambda \cdot v_i(t) \tag{33}$$

Hence:

$$1 = \frac{\partial \ln(v_i)}{\partial \ln \lambda} = \frac{\partial \ln(v_i)}{\partial \ln \lambda_t} \cdot \frac{d \ln \lambda_t}{d \ln \lambda} + \sum_{j=1}^{n} \frac{\partial \ln(v_i)}{\partial \ln \lambda_j} \cdot \frac{d \ln \lambda_j}{d \ln \lambda} = -C_t^{v_i} + \sum_{j=1}^{n} C_j^{v_i} \tag{34}$$

At steady state this leads to the standard law for flux control coefficients, i.e. that their sum must equal 1.

The concentration control coefficients for dynamic processes determined at a fixed time point, such as used above in this manuscript and also by [23] can become extremely large for periodic processes, where parameters may change the frequency, as well as in very slowly relaxing systems. The concept of 'progression concentration control coefficient', which is determined at a fixed 'phase' in a dynamic response has been introduced to deal with this problem [8]. In the progression control coefficient the perturbation of the 'phase' of the response is corrected for by determination of the control coefficients at a given progress point (corresponding to the phase in a periodic phenomenon). That perturbation of phase is the essence of the time control coefficients. The relation between the two types of control coefficients can be described as:

$$C_j^{c_i(t)} = C_j^{c_i(pA)} - C_t^{c_i(t)} C_j^A \tag{35}$$

where the control by reaction j on the concentration c_i at a fixed time point t is expressed as a function of the concentration control by j at a point in the progress curve corresponding to t (i.e. pA) and the control by time on the concentration, multiplied by the control of reaction j on the length of the progress curve (A). The latter control measures the extent to which the modulation of reaction j stretches (retards) the time dependence of the process.

The analysis is valid for any type of dynamic behavior and the point A can be set as a threshold of a concentration but can also be the maximum of a periodic signal. One of the advantages of using the progression control analysis is that it is not sensitive to the growing dispersion between the original and the perturbed signal, which is especially important for analyses over a prolonged time period, for instance when analyzing a periodic signal.

A summation of equation 1 over all the reaction steps yields:

$$\sum_{j=1}^{n} C_j^{c_i(t)} = \sum_{j=1}^{n} C_j^{c_i(pA)} - C_t^{c_i(t)} \sum_{j=1}^{n} C_j^A \tag{36}$$

with

$$\sum_{j=1}^{n} C_j^A = -1 \tag{37}$$

and

$$\sum_{j=1}^{n} C_j^{c_i(pA)} = 0 \tag{38}$$

which reduce Eq. 2 to Eq. 32. This shows that correction for the 'phase-shift' renders the classical summation theorem.

This flux control summation law for steady state (Eq. 34 with time control zero) shows that in a network of 14 processes, the average flux control coefficient of the enzyme is 1/14, i.e. only some 7%, i.e. most of its flux is determined by the other enzymes. By implementing its connectivity laws, or the matrix inversion methodology [43], Metabolic Control Analysis also enables to determine the actual (rather than pathway-average) control the enzyme has over its own flux, and why this control is as high as it is, but we shall not further detail this in the present paper. With this, Metabolic Control Analysis fulfils one promise of Systems Biology, i.e. to help understand how collectively molecular interaction properties lead to system properties.

The other fundamental aspect relates to the science of non-equilibrium thermodynamics. This science has discovered some general principles that should apply to systems that are in steady states that are close to equilibrium. However, Biological systems are far enough from equilibrium for the near-equilibrium assumptions not to apply. For quite a while there has been only slow progress in devising a non-equilibrium thermodynamics that would apply profitably to living systems. An exception has been Mosaic Non-Equilibrium Thermodynamics, which broke with the tradition of thermodynamics of only being interested in completely general properties [42].

As shown above, MCA is based on recognizing that the basis of all biological organization resides in a limited number of ground rules, as epitomized in Eq. 5. These ground rules may be treated as postulates for which ample experimental evidence is available. With the limitation of operating within systems that adhere to the postulates, MCA then reveals a number of general principles of operation such as the summation law. This situation is fairly similar to that of statistical thermodynamics and the second law of thermodynamics, which can also be derived from a set of basic postulates to which most actual systems adhere. As MCA deals with non-equilibrium systems, it is perhaps the extension of Non-Equilibrium Thermodynamics we have long been looking for.

Because in its early years many theoretical examples and experimental applications of MCA dealt with ideal linear metabolic pathways at steady state, many scientists have the impression that it is limited to those pathways. This is not the case [13]. The summation laws apply to any metabolic network with the sometimes confusing or inspiring effect that control of a flux through a pathway may well lie in the network outside that pathway [21]. The laws also apply to pathways that are not ideal in the sense that more than one enzyme is involved in a single reaction [23] or metabolites are channeled between enzymes [27]. MCA also applies to signal transduction and gene expression pathways where it is indeed worthwhile further to extend the analysis to Hierarchical Control Analysis [25,39]. And, as shown for instance by [8,11,16–18] and Eqs. 32 and 34d, MCA is not confined to steady states either; oscillations and transients can also be dealt with.

4.1.3.3 Example: control of nuclear receptor signaling by time, in silico In a number of cellular phenomena time may be an important factor. Then the magnitude of the control by time, as quantified by the corresponding control coefficient (Eq. 27), is of interest. We emphasize that we are here not discussing the evolution in time of the control of features of the system by various molecular activities, by the control by time

itself, which is a novelty for Metabolic Control Analysis. For instance, if the control by time is doubled at all time points, this means that the entire process progresses twice as fast, and that the total activation of transcription may be halved. In a different scenario, cell function may be determined by two processes, such as growth factor induced transcription activation and the cell cycle. Control by time of the former process would imply control of the relative phases of the two processes and hence control of the decision between differentiation and proliferation.

Nuclear Receptors (NRs) are involved in the modulation of intra and extra cellular signals into the realization of specific gene programs [5]. The mechanism of transcription regulation varies between different groups of NRs. Some of them are largely located in the nucleus (vitamin D receptors) and after addition of the ligand switch their activity from being inhibitors to becoming activators of transcription [32]. Others, like glucocorticoid receptors (GRs) are largely located in the cytoplasm [49]. Upon activation by ligand the latter move to the nucleus and initiate the transcription of responsive genes. The second mechanism, which potentially also depends on the ligand-dependent activation of the nucleo-cytoplasmic (n/c) transport of nuclear receptors, is of particular interest [2,31]. On the one hand, the additional check-points of signal transduction provided (i.e. at n/c transport) may safeguard signaling from the occasional error or noise. On the other hand, it offers additional possibilities to regulate signal transduction by other cellular processes.

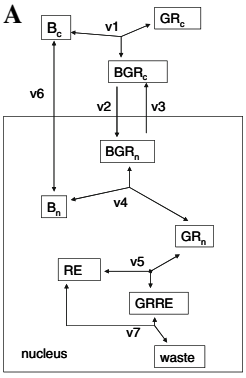
GR signaling plays an important regulatory role in gluconeogenesis and glucose uptake, lipolysis in adipose tissues, proteolysis in muscles, osteoblast differentiation and apoptosis [12,50]. Taking into account the vital function of these processes we can understand why GR signaling employs the second mechanism based on the ligand-dependent activation of GR n/c transport.

We are studying the GR signaling system by use of a simplified model, characterized by the following considerations (Fig. 1):

- In the cytoplasm, active GR_c binds with molecules out of the cytoplasmic pool of importins (B_c) in a reversible MA (mass action kinetics) reaction (v_1).
- The resultant BGR_c complex is transported into the nucleus described by an irreversible MM (Michaelis–Menten kinetics) reaction (v_2) and returns back from the nucleus to cytoplasm in an irreversible MM reaction (v_3) (cf. Eq. 15 with V_{-k} zero).
- In the nucleus the complex of active GR-importins (BGR_n) splits into active GR_n and nuclear importin (B_n) in a reversible MA reaction (v_4)
- Active GR interacts with its RE (part of the chromosomal DNA) by a MA reaction (v_5)
- Importin is transported back to the cytoplasm in a reversible MA reaction (v_6)
- GR when bound to RE is degraded in the irreversible MA reaction (v_7).

Figure 2 shows that, according to the model, addition of a certain amount of mobile Glucocorticoid Receptor (GRc) results in an increase of the concentration of GRRE. If there is no degradation of GRRE complex ($v_7 = 0$), the concentration of the latter relaxes monotonically to an equilibrium relationship with GR_n (Fig. 2a). If v_7 is not equal to 0, we observe a peak in GRRE concentration. After the peak, the concentration of GRRE returns to 0 (Fig. 2b).

A



B

	Reactions	parameters
$\mathbf{V_1}$	Reversible Mass Action: $k_{1f} \cdot B_c(t) \cdot GR_c(t) - k_{1b} \cdot B_c GR_c(t)$	K_{1f}=4; k_{1b}=1
$\mathbf{V_2}$	Irreversible Michaelis-Menten: $\frac{V_{m2} \cdot BGR_c(t)}{K_{m2} \cdot (K_{m2} + BGR_c(t))}$	V_{m2}=4; K_{m2}=0.05
$\mathbf{V_3}$	Irreversible Michaelis-Menten: $\frac{V_{m3} \cdot BGR_n(t)}{K_{m3} \cdot (K_{m3} + BGR_n(t))}$	V_{m3}=20; K_{m3}=0.5
$\mathbf{V_4}$	Reversible Mass Action: $k_{4f} \cdot BGR_n(t) - k_{4b} \cdot B_n(t) \cdot GR_n(t)$	k_{4f}=200; k_{4b}=20
$\mathbf{V_5}$	Reversible Mass Action: $k_{5f} \cdot GR_n(t) \cdot RE(t) - k_{5b} \cdot GRRE(t)$	k_{5f}=20; k_{5b}=2
$\mathbf{V_6}$	Reversible Mass Action: $k_{6f} \cdot B_n(t) - k_{6b} \cdot B_c(t)$	k_{6f}=1; k_{6b}=1
$\mathbf{V_7}$	Irreversible Mass Action: $k_7 \cdot GRRE(t)$	k_7=25
Conservation Moieties: $B_c(t)+B_n(t)+BGR_c(t)+BGR_n(t)=B_{total}$ $GR_c(t)+BGR_c(t)+BGR_n(t)+GR_n(t)+GRRE(t)+waste=GR_{total}$ $RE(t)+GRRE(t)=RE_{total}$		B_{total}=15 GR_{total}=1 RE_{total}=1
Initial conditions: $GR_c(0)=GR_{total}$; $BGR_c(0)=0$; $BGR_n(0)=0$; $GR_n(0)=0$; $GRRE(0)=0$; $waste(0)=0$; $RE(0)=RE_{total}$; $B_n(0)$ and $B_c(0)$ are in thermodynamic equilibrium ($B_n(0) = \frac{bt \cdot \frac{k6b}{k6f}}{1+\frac{k6b}{k6f}}$; $B_c(0) = \frac{bt}{1+\frac{k6b}{k6f}}$)		

Fig. 1 Simplified GR signaling model: **a** network structure, **b** equations and kinetic parameters used. Note that both the model and parameter values do not correspond to reality. This model serves only illustrative purposes

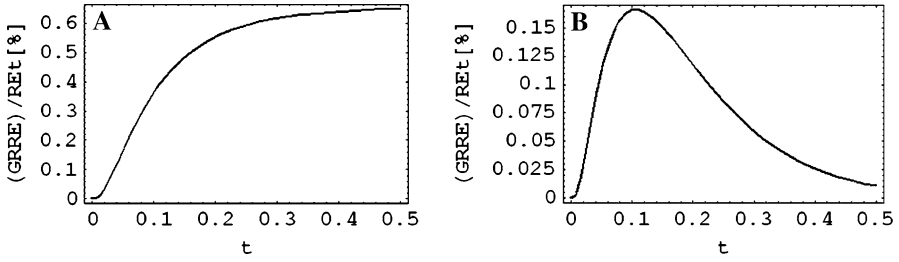


Fig. 2 Change in the concentration of GRRE as a function of time in GR signaling. **a** as calculated by the simplified model (Fig. 1), lacking degradation of GRRE complex ($k_7 = 0$); **b** with degradation of the GRRE complex ($k_7 = 25$)

Table 1 Calculated GRRE concentration control coefficients (by individual reactions and by time)

Control	Time			
	$t = 0.07$	$t = 0.105$	$t = 0.2$	$t = 0.5$
C_1^{GRRE}	0.383	0.189	−0.136	−0.417
C_2^{GRRE}	0.030	0.012	−0.008	−0.068
C_3^{GRRE}	−0.019	−0.012	0.000	0.056
C_4^{GRRE}	0.039	0.012	0.015	−0.073
C_5^{GRRE}	0.625	0.473	0.064	−1.989
C_6^{GRRE}	0.002	0.003	0.004	0.025
C_7^{GRRE}	−0.480	−0.669	−1.082	−2.012
$\sum_{i=1}^{7} C_i^{\mathrm{GRRE}}$	0.580594	0.007078	−1.17309	−4.52682
C_t^{GRRE}	0.580594	0.007078	−1.17309	−4.52682

For the determination of the control by time we calculated the time derivative of each of the variables and normalized these at the respective time points (i.e. $CXt = (dX/dt \times t/X)$)

We estimated the control of GRRE concentration by time, for example at $t = 0.07$, as follows (Table 1):

$$C_t^{\mathrm{GRRE}} \approx \frac{\Delta \ln[GRRE]}{\Delta \ln t} = 0.58$$

Subsequently, we examined how the level of GRRE was controlled by the individual reactions at that particular time ($t = 0.07$). For example, for the reaction of export of importin (v_6) we calculated $C_{\mathrm{v6}}^{\mathrm{GRRE}}$ as the change in GRRE concentration resulting from a small perturbation in the constant k_6. $C_{\mathrm{v6}}^{\mathrm{GRRE}}$ equaled 0.002. We performed the same calculations for all reactions included in GR signaling model and found all control coefficients and their sum $\sum_{i=1}^{n} C_{vi}^{\mathrm{GRRE}}$ (Table 1). $\sum_{i=1}^{n} C_{vi}^{\mathrm{GRRE}}$ equaled 0,58, which is the same as the control by time ($C_t^{\mathrm{GRRE}} = 0.58$). In this way we validated and illustrated the summation law of Eq. 32.

We found that in the first phase (e.g. at $t = 0.07$) the concentration of active nuclear receptor increased with time as reflected by the control by time being positive. Table 1 indicates the processes that exercised positive control on this concentration and the processes that exerted negative control. The fact that the control by time was positive, inferred that the positively controlling processes were more important for the magnitude of GRRE than the negatively controlling processes were. If one looks for possible therapeutics this might be helpful.

However at $t = 0.105$, i.e. at the peak of the GRRE concentration, the sum of all positive controls ($\sum +C_k^{\mathrm{GRRE}}$) became equal to the sum of all negative controls ($\sum -C_k^{\mathrm{GRRE}}$) and thereafter, with further increase of time, negative control coefficients started playing the major role (Table 1).

As has been stated before, Metabolic Control Analysis (MCA) can be considered as a conceptual way of describing system behavior by simplifying and limiting questions asked. One type of limiting question we have asked is how particular systemic

behavior is controlled by one particular reaction, for example how the GRRE concentration is controlled by the reaction of the export of importins (which is quantified as the control coefficient C_{v6}^{GRRE}). This control coefficient depends on environmental conditions, structure of biochemical network, rates of reactions and, as we have shown, it depends on time. The next question is to measure how this control coefficient is controlled by all these factors at different times. For our particular case the answer addresses a very practical issue—how to make GR signaling to be controllable (or not controllable) by the reaction of export of importins? At what time will the signaling be the most (least) controllable? The approach of the simplification via MCA should enable one to address these questions effectively and to link particular design features and conditions of the system with its systemic properties. The understanding of this link implies understanding of design principles of the studied system.

4.2 Inverse modeling

Inverse modeling ranges between methods that generate phenomenological models that then predict (or rather 'postdict') the observed functional behavior of the system, to models that generate precise mechanistic models inclusive of values for parameters that have mechanistic meaning. Indeed, a number of the approaches discussed above can also be used in inverse modeling. One approach is to scan parameter values and for each parameter value calculate system performance in a silicon cell type of approach. Another one uses metabolic control analysis in an inverse mode, i.e. a matrix of control coefficients is inverted to calculate the elasticity coefficients required to exhibit such a control distribution [42,46].

A new method (Westerhoff et al., in preparation) measures the metabolite concentrations experimentally and inserts these into the rate equations. It then uses flux analysis to calculate the fluxes through all the reactions. Subsequently it tunes the rates to the known rates by adjusting the concentrations of the enzymes. Provided all experimental free-energy differences across the reactions have the sign that corresponds with the direction of the flux through the reaction, this should always lead to a solution. As most reaction rates depend proportionally on (functions of the) concentration of the enzyme that catalyzes them) this is a rather simple procedure.

The former type of inverse modeling has the attractiveness that it is hypothesis free, and indeed corresponds to data-driven hypothesis generation. Examples are genetic networks constructed on the basis of observed behavior of mRNA concentrations [10]. A disadvantage of this method is that it disregards what is already known, even aspects that are known for sure, e.g. that mRNA's do not directly affect the transcription of genes but require the protein and metabolite parts of the cellular organization to mediate the regulation. Inverse modeling approaches such as naïve versions of flux balance analysis may generate models that violate the second law of thermodynamics.

The latter type of inverse modeling is also problematic, as it is not absolutely clear for each bit of existing knowledge how solid it is, and because models based on existing knowledge may require too many parameters to be fitted in the inverse procedure. The models are 'underdetermined'; many combinations of parameters fit the experimental data. In practice then, the latter type of inverse modeling is a mix

between reverse and forward modeling, as part of what is already known about the system (such as the network topology) is considered fixed and only some aspects are induced [3,4] from the experimentally observed behavior. Indeed, the key strength of this approach to inverse modeling is that it works with mechanistically valid models and allows existing biochemical knowledge (prior knowledge in the Bayesian sense) to be incorporated into the estimation process. Recent approaches have formalized this trade-off between prior kinetic knowledge and the agreement with experimental data into effective algorithmic frameworks [47].

Since each of the general approaches outlined above has its own merits and can be seen as complimentary, a sophisticated mix between the two may be best.

5 Concluding remarks

In this paper we have given an overview of the part of the new systems biology that relates to metabolic and signal transduction pathways. We have shown that it extends much of existing mathematical biochemistry. With the silicon cell approach and Metabolic and Hierarchical Control Analysis it is able to highlight where and how system function derives from molecular interactions. The new law that relates concentration control to control by time was illustrated for an important signal transduction pathway. It is envisaged that there is much more Mathematical Biology to be discovered in this area between molecules and Life.

Acknowledgments This paper is dedicated to the memory of Reinhart Heinrich, who was a teacher and friend but above all an example to HVW and to many others. We thank various research funding organizations for support, including NBN (South Africa) for the support of RC and the JWS Online project, EPSRC, BBSRC, NWO, EZ-NL, and EC-FP6 (BioSim, NucSys, YSBN, EC-MOAN). Indeed NWO, NIH, and BBSRC should be acknowledged for helping fundamental biochemistry and Metabolic Control Analysis survive the attrition years of the seventies and eighties. It is only in the present century that their role as pillars under Systems Biology is becoming appreciated.

References

1. Alberghina, L., Westerhoff, H.V. (eds.): Systems Biology: Definitions and Perspectives, 408 pages. Springer, Berlin, ISBN 354022968X (2005)
2. Albermann, L., Shahin, V., Ludwig, Y., Schafer, C., Schillers, H., Oberleithner, H.: Evidence for importin-alpha independent nuclear translocation of glucocorticoid receptors in Xenopus laevis oocytes. Cell. Physiol. Biochem. **14**, 343–350 (2004)
3. Boogerd, F.C., Bruggeman, F., Hofmeyr, J.H.S., Westerhoff, H.V. (eds.): Systems Biology Philosophical Foundations, 1st edn, 342 pages. Elsevier, Amsterdam, ISBN: 0444520856 (2007)
4. Bruggeman, F.J., Westerhoff, H.V.: The nature of systems biology. Trends Microbiol. **15**, 45–50 (2007)
5. Carlberg, C., Dunlop, T.W.: An integrated biological approach to nuclear receptor signaling in physiological control and disease. Crit. Rev. Eukaryot. Gene Expr. **16**, 1–22 (2006)
6. Chance, B., Williams, G.R., Holmes, W.F., Higgins, J.: Respiratory enzymes in oxidative phosphorylation. 5. A mechanism for oxidative phosphorylation. J. Biol. Chem. **217**, 439–451 (1955)
7. Chen, K.C., Calzone, L., Csikasz-Nagy, A., Cross, F.R., Novak, B., Tyson, J.J.: Integrative analysis of cell cycle control in budding yeast. Mol. Biol. Cell **15**, 3841–3862 (2004)
8. Conradie, R., Westerhoff, H.V., Rohwer, J.M., Hofmeyr, J.-H.S., Snoep, J.L.: Summation theorems for flux and concentration control coefficients of dynamic systems. IEE Proc. Syst. Biol. **153**, 314–317 (2006)

9. Cornish-Bowden, A.: Fundamentals of Enzyme Kinetics, 3rd edn., 344 pages. Portland Press, London, ISBN 1 85578 1581 (2004)
10. Dasika, M., Gupta, A., Maranas, C.D., Varner, J.D.: A mixed integer linear programming (MILP) framework for inferring time delay in gene regulatory networks. Pac. Symp. Biocomput. **9**, 474–485 (2004)
11. Demin, O.V., Westerhoff, H.V., Kholodenko, B.N.: Control analysis of stationary forced oscillations. J. Phys. Chem. B **103**, 10695–10710 (1999)
12. Eijken, M., Koedam, M., et al.: The essential role of glucocorticoids for proper human osteoblast differentiation and matrix mineralization. Mol. Cell. Endocrinol. **248**(1–2), 87–93 (2006)
13. Fell, D.A.: Increasing the flux in metabolic pathways: a metabolic control analysis perspective. Biotechnol. Bioeng. **58**, 121–124 (1998)
14. Garfinkel, D., Hess, B.: Metabolic Control Mechanisms. Vii. A detailed computer model of the glycolytic pathway in ascites cells. J. Biol. Chem. **239**, 971–983 (1964)
15. Goldbeter, A.: Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behavior, 629 pages. Cambridge University Press, Cambridge, ISBN 0521403073 (1996)
16. Heinrich, R., Rapoport, T.A.: Mathematical analysis of multienzyme systems. II. Steady-state and transient control. BioSystems **7**, 130–136 (1975)
17. Heinrich, R., Reder, C.: Control analysis of relaxation processes. J. Theor. Biol. **151**, 343–350 (1991)
18. Heinrich, R., Neel, B.G., Rapoport, T.A.: Mathematical models of protein kinase signal transduction. Mol. Cell **9**, 957–970 (2002)
19. Heinrich, R., Rapoport, T.A.: Linear steady-state treatment of enzymatic chains—general properties, control and effector strength. Eur. J. Biochem. **42**, 89–95 (1974)
20. Heinrich, R., Rapoport, S.M., Rapoport, T.A.: Metabolic regulation and mathematical models. Prog. Biophys. Mol. Biol. **32**, 1–82 (1977)
21. Hoefnagel, M.H.N., Starrenburg, M.J.C., Martens, D.E., Hugenholtz, J., Kleerebezem, M., Van Swam, I.I., Bongers, R., Westerhoff, H.V., Snoep, J.L.: Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modeling, metabolic control and experimental analysis. Microbiology-Sgm **148**, 1003–1013 (2002)
22. Holzhütter, H.G.: The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur. J. Biochem. **271**, 2905–2922 (2004)
23. Hornberg, J.J., Bruggeman, F.J., Binder, B., Geest, C.R., de Vaate, A., Lankelma, J., Heinrich, R., Westerhoff, H.V.: Principles behind the multifarious control of signal transduction—ERK phosphorylation and kinase/phosphatase control. FEBS J. **272**, 244–258 (2005)
24. Kacser, H., Burns, J.A.: The control of flux. Symp. Soc. Exp. Biol **27**, 65–104 (1973)
25. Kahn, D., Westerhoff, H.V.: Control-theory of regulatory cascades. J. Theor. Biol. **153**, 255–285 (1991)
26. Katchalsky, A., Curran, P.F.: Nonequilibrium Thermodynamics in Biophysics, 248 pages. Harvard University Press, Cambridge, ISBN:0674625501 (1965)
27. Kholodenko, B.N., Cascante, M., Westerhoff, H.V.: Control-theory of metabolic channeling. Mol. Cell. Biochem. **143**, 151–168 (1995)
28. Lawrence, P.A.: The Making of a Fly: The Genetics of Animal Design, 240 pages. Blackwell, Oxford, ISBN: 978-0632030484 (1992)
29. Meinhardt, H.: Models of Biological Pattern Formation, 230 pages. Academic Press, New York, ISBN:978-0124886209 (1982)
30. Olivier, B.G., Snoep, J.L.: Web-based kinetic modelling using JWS Online. Bioinformatics **20**, 2143–2144 (2004)
31. Pemberton, L.F., Paschal, B.M.: Mechanisms of receptor-mediated nuclear import and nuclear export. Traffic **6**, 187–198 (2005)
32. Pike, J.W., Glorieux, F.H., Feldman, D.: Vitamin D, 2nd edn., 1952 pages. Academic Press, New York, ISBN13: 9780122526879 (2005)
33. Price, N.D., Papin, J.A., Schilling, C.H., Palsson, B.O.: Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol. **21**, 162–168 (2003)
34. Rapoport, I., Berger, H., Elsner, R., Rapoport, S.M.: Eur. J. Biochem. **73**, 421–427 (1977)
35. Rapoport, T.A., Rapoport, S.M., Reich, J.G., Hohne, W.E., Heitmann, P.: Kinetic model for action of inorganic pyrophosphatase from bakers-yeast—activating influence of magnesium ions. Eur. J. Biochem. **26**, 237–246 (1972)
36. Reijenga, K.A., Westerhoff, H.V., Kholodenko, B.N., Snoep, J.L.: Control analysis for autonomously oscillating biochemical networks. Biophys. J. **82**, 99–108 (2002)

37. Savageau, M.A.: Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology, 199 pages. Addison-Wesley, New York, ISBN:978-0201067385 (1976)
38. Snoep, J.L., Bruggeman, F., Olivier, B.G., Westerhoff, H.V.: Towards building the silicon cell: a modular approach. Biosystems **83**, 207–216 (2006)
39. Snoep, J.L., van der Weijden, C.C., Andersen, H.W., Westerhoff, H.V., Jensen, P.R.: 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 (2002)
40. Van der Meer, R., Westerhoff, H.V., Van Dam, K.: Linear relation between rate and thermodynamic force in enzyme-catalyzed reactions. Biochim. Biophys. Acta **591**, 488–493 (1980)
41. Westerhoff, H.V.: Signalling control strength. J. Theor. Biol., in press (2007). doi:10.1016/j.jtbi.2007.11.035
42. Westerhoff, H.V., Kell, D.B.: Matrix method for determining steps most rate-limiting to metabolic fluxes in biotechnological processes. Biotechnol. Bioeng. **30**, 101–107 (1987)
43. Westerhoff, H.V., Kell, D.B.: What biotechnologists knew all along...? J. Theor. Biol. **182**, 411–420 (1996)
44. Westerhoff, H.V., Palsson, B.O.: The evolution of molecular biology into systems biology. Nature Biotechnol. **22**, 1249–1252 (2004)
45. Westerhoff, H.V., Van Dam, K.: Thermodynamics and Control of Biological Free-Energy Transduction, 568 pages. Elsevier, Amsterdam, ISBN: 0-444-80783 (1987)
46. Westerhoff, H.V., Hofmeyr, J.H.S., Kholodenko, B.N.: Getting to the inside of cells using metabolic control analysis. Biophys. Chem. **50**, 273–283 (1994)
47. Wilkinson, S.J., Benson, N., Kell, D.B.: Proximate parameter tuning for biochemical networks with uncertain kinetic parameters. Mol. BioSystems, DOI:10.1039/ B707506E (2007)
48. Wu, L., Wang, W.M., van Winden, W.A., van Gulik, W.M., Heijnen, J.J.: A new framework for the estimation of control parameters in metabolic pathways using lin–log kinetics. Eur. J. Biochem. **271**, 3348–3359 (2004)
49. Yang, J., DeFranco, D.B.: Assessment of glucocorticoid receptor heat shock protein 90 interactions in vivo during nucleocytoplasmic trafficking. Mol. Endocrinol. **10**, 3–13 (1996)
50. Zhou, J.G., Cidlowski, J.A.: The human glucocorticoid receptor: One gene, multiple proteins and diverse responses. Steroids **70**, 407–417 (2005)