

# What is systems biology? From genes to function and back

Hans V. Westerhoff and Jan-Hendrik S. Hofmeyr

## Abstract

The essence of the grand contributions of physiology and molecular biology to biology is discussed in relation to what may be needed to understand living systems. Unanswered is the link between function and molecular behaviour, and emergence of function from the nonlinear interactions, respectively. Systems biology should focus on properties that emerge in nonlinear interactions from the molecular level up, which are crucial for biological function. Pre-genomics approaches such as Metabolic and Hierarchical Control Analysis have already contributed concepts and conclusions to systems biology. Their combination with the genome-wide analyses should now lead to substantial progress in the understanding of life. An aspect of biology at odds with traditional physics and chemistry is the circular causation that occurs in all living systems. By analyzing this phenomenon quantitatively, systems biology can already deal with certain types of circular causation by dissection.

## 1 What came before?

### 1.1. Physiology

The experts differ on whether systems biology has been around for a while or if it is a relatively new science. Both sides may be right, as we shall argue in this chapter. Perhaps biology started in earnest when human beings marvelled over spermatozoa as seen under a microscope, trying to recognize the *homunculus* (little man) in them (*i.e.* supposing that a complete system should be there). Or perhaps biology started with the study of human anatomy, where scientists and artists alike marvelled at the high level of organization in terms of well-defined and fairly autonomous organs, with functions that could almost be understood. Although they were studying biological systems, these disciplines have been called physiology rather than systems biology. They engage in *discourses of nature* where the word ‘discourse’ is significant in the sense of often being argumentation-driven rather than just data-driven.

Physiology is attractive precisely because it relates directly to function. Indeed, should function be understood, then dysfunction should be understood as well, and avenues towards the treatment of disease should open up. The *discourses* of

physiology were often based on rather loose principles however, which appeared to lack generality and were always open to *ad hoc* exceptions. If physiology was at all based on 'laws' of nature, these laws tended to be empirical, such as the law of Haeckel (phylogeny being a recapitulation of ontogeny) and the scale-free relation between respiration rate and biomass. This physiology was often little more than descriptive, useful nonetheless when features would be recurrent. After all, proper diagnosis and then treatment based on experience with previous cases is still one of the most successful qualities of human medicine.

For some time, it seemed that the laws or principles of physiology might be akin to the early formulations of the second law of thermodynamics. This law started off as an empirical law, which appeared to be liable to falsification in any new system under study. The second law of thermodynamics gained enormous standing when it could be derived from the realization that much matter consists of large numbers of rather ill-organized particles, which due to their randomness caused systems to produce entropy, provided that entropy was redefined in terms of an average of molecular properties (*cf.* Westerhoff and Van Dam 1987).

## 1.2 Molecular biology

A remarkable event occurred when Mendel observed regularities when crossing plants, which could be deduced to very simple rules governing the behaviour of quasi particles, later called genes. Apparently, the appearance of a plant system was determined by underlying agents (possibly material). Later, the discovery of the DNA structure with its facile explanation of much of genetics, provided more of a basis to this, however, by then the real biological revolution had already taken place.

The biological revolution had been preceded by the chemical revolution where it had been recognized that the dead world of physics was particle-based. Essentially, this referred to a quantum nature of matter, in the sense that matter comes in different types, all with different properties. These properties are discontinuous, *i.e.* gold differs from silver and there is not necessarily matter with properties halfway in-between. For instance, when hydrogen is made to react with oxygen, the result has completely new properties. The concepts of atoms and molecules revolutionized the ways in which one analyzed the world and for many years the chemical industry delivered many new materials with new and useful properties. Chemistry was perhaps the first clear systems science.

It was soon observed that many of the same molecules that occurred in inorganic matter also occurred in living systems. The new science of chemistry had many ties with biology. No chemical elements were discovered that were unique to living matter. And in fact many chemical molecules for which properties and structure were determined *ex vivo* had a biological origin. Boiling living matter in hydrogen chloride frees a large number of small molecules, including nucleosides and amino acids, all of which inspired organic chemistry into making more similar molecules that could be useful to mankind.

This phase in scientific history suggested that life was perhaps little more than a collection of such molecules. It also inspired *biochemistry* into a search for the reaction pathways through which those molecules were synthesized. And indeed, rapid progress ensued, up to the elucidation of the molecular basis of life. A major step was the recognition that virtually every chemical reaction in living organisms was catalyzed by a protein and that, therefore, the chemical pathways in life could be delineated by isolating and characterizing the proteins that catalyzed the subsequent steps. Next steps were the discovery of a correspondence between those very same processes and the genes discovered by genetics, and the subsequent identification of genes with parts of the long linear information carrying molecules of DNA. The connection between genetics and biochemistry led to the recognition that life could be studied successfully at the molecular level. This was expressed in the term ‘molecular biology’, which had an emphasis on the principle that DNA is expressed through RNA into protein, which then catalyzes molecular processes. The primary structure of genes could be elucidated, the corresponding amino acid sequence deduced, the 3-dimensional structure of the corresponding protein could be determined, and action and the mechanism of action of many proteins could be established. Indeed, of every macromolecule, and of every process in living organisms everything could be determined and explained, or so it seemed. This became the triumph of molecular biology and biochemistry combined.

### 1.3 Systems molecular biology?

So, here we are. We will (soon) be able to determine the identity and concentrations of molecules in living organisms. We must surely be close to understanding life and curing its diseases!

While physiology had come close to describing function without really understanding it in solid physical chemical terms, molecular biology now seemed to understand life in precisely those terms, or, at least to do so for the molecules in life. There was an issue with molecular biology, as with physiology, that it might remain an incomplete science and, therefore, it remained a limited scientific discipline in most of the previous century. Living organisms are specified by so many genes and proteins that it seemed that molecular biology could never fully characterize a living system. As a consequence, demonstrated failures of molecular biology to understand living systems could always be attributed to the existence of still unknown factors.

This is where *genomics* caused yet another revolution. With the sequencing of entire genomes and with the ability to study their expression at the level of transcriptome, proteome, and metabolome, every molecular factor became identifiable. And indeed, much activity went into complete identification of transcriptomes, proteomes, and metabolomes. Complete inventories of systems were and are made. Molecular biology became a complete science; no elusive factors remained and scientific explanations by molecular biology became falsifiable, or so it seemed.

Some equated this new molecular biology with systems biology, as it comprised the complete molecular identification of the system of any living organism. With the success of molecular biology in helping us to understand the mechanisms of the individual molecules of life, the extension of molecular biology to the complete set of molecules of life would surely help us comprehend life. Understanding 30,000 times 1/30,000th part of life would surely amount to the understanding of life itself or would it? Is systems biology nothing more than molecular biology of entire systems? Should it be named ‘systems molecular biology’ (Westerhoff and Palsson 2004)?

## 2 Limits to systems molecular biology

### 2.1 Data floods

A number of caveats with respect to such optimism soon emerged. One of these might prove to be technical only: there is too much data. Data on transcriptomics, proteomics, and metabolomics are now being accumulated at higher rates than that they can be analyzed and structured.

Bioinformatics comes to the rescue here, and so is the ever-increasing power of digital computers. Moreover, the number of types of molecules is limited and for understanding the essence of life, we need perhaps not understand all possible conditions. Furthermore, a better specification of the biological issue one actually wishes to address, and a return to hypothesis-driven research might limit the amount of data needed for analysis significantly. Therefore, molecular biology, genomics plus bioinformatics might still be able to do the job.

### 2.2 Nonlinearity

A second caveat relates to something that is already well known: much of biological function stems from inherent strong nonlinearity. The word nonlinearity is here used in a broad sense and deserves appreciable specification. We shall make this specification by the use of some simple algebra. We ask the less mathematically oriented reader to bear with us; it will be well worth to study this example, because it gives the *crux* of why systems biology is more than the new clothes of the emperor!

We consider molecules of type  $x$  and molecules of type  $y$  and first assume that a functional property of interest,  $f$ , depends linearly on both  $x$  and  $y$ :

$$f = a \cdot x + b \cdot y + c \quad (1)$$

$x$  could be the number of molecules of  $X$ ,  $y$  the number of molecules of  $Y$ ,  $a$  and  $b$  their respective molecular masses and  $c$  the mass of the rest of the cell;  $f$  being the mass of the total. The linear dependence of  $f$  on both  $x$  and  $y$  independently has the important property that one can study the dependence of  $f$  on  $X$  and  $Y$  independently and then understand how  $f$  behaves by just superimposing the dependencies.

For, denoting changes by delta's, one can first determine experimentally how function changes when only  $x$  changes and, thus, determine the value of  $a$ .

$$\frac{\Delta f}{\Delta x} = a \quad (2)$$

Similarly one can determine the value of  $b$ :

$$\frac{\Delta f}{\Delta y} = b \quad (3)$$

When both  $x$  and  $y$  change one can then simply calculate the change in function from  $a$ ,  $b$  and the actual changes in  $x$  and  $y$ , through:

$$\Delta f = a \cdot \Delta x + b \cdot \Delta y \quad (4)$$

For the mass of a dead cell this might work. But it may not work for many other functions. For instance, it would not work for the mass of a living cell. For, the change in amount of a certain enzyme  $x$  would lead to a change in the rate of the reaction it catalyzes and, therefore, to a change in metabolite concentrations and because the cell is an open system a corresponding change in total mass. The latter change could well depend on whether or not the amount of a second enzyme  $y$  is also changed at the same time. In the latter case, the total change in mass should not equal the change in mass of enzyme  $x$  plus the change in mass of enzyme  $y$ .

Looking at the living cell from a structural point of view, one might, inadvertently, come to the view that total structure is a linear property: determine the structure of all proteins independently and call these structures  $x$ ,  $y$ , etc., multiply each structure with the number of molecules of the corresponding protein ( $a$ ,  $b$ , etc.). Do this for all proteins and add up all the results so as to obtain the total structure of the cell (one should of course add spatial coordinates for where in the cell the structure sits). The approach seems reasonable, and is valuable perhaps as a first order approach. Yet, one quickly realizes that it may still be incorrect even if one only focuses on the structure and not on the dynamics and functioning of the cell. The approach should fail for instance if two proteins (*e.g.* subunits of a single protein) interact with each and form a complex of a more compact structure, or if chaperons are involved, or if they interact with each other and cause the synthesis of ATP, which causes the phosphorylation of a kinase, the expression of many genes and, therefore, altered levels of many other proteins.

Let us now look at a case where there is no linear relationship between function and the molecules  $x$  and  $y$ , *e.g.*:

$$f^* = d \cdot x \cdot y^2 \quad (5)$$

Even though this equation is even simpler than the linear one (fewer parameters, *i.e.* only  $d$  rather than  $a$ ,  $b$ , and  $c$ ), it produces many complications. First, the dependence of function ( $f^*$ ) on  $x$  is no longer constant but depends on  $y$ :

$$\frac{\Delta f^*}{\Delta x} = d \cdot y^2 \quad (6)$$

and, hence, cannot be determined once and for all: The functioning of  $x$  depends on the activity of  $y$ .

For the dependence of function on  $y$  the situation is even worse: it depends on  $x$ , on  $y$ , as well as upon the change in  $y$ :

$$\frac{\Delta f^*}{\Delta y} = d \cdot x \cdot (y + \Delta y)^2 - d \cdot x \cdot y^2 = 2 \cdot d \cdot x \cdot y + d \cdot x \cdot \Delta y \quad (7)$$

In addition, a change in function due to a change in both  $x$  and  $y$  cannot be understood as the sum of the change in function due to the change in  $x$  and the change in function due to the change in  $y$ :

$$\begin{aligned} \Delta f^* &= 2 \cdot d \cdot x \cdot y \cdot \Delta y + d \cdot x \cdot (\Delta y)^2 + d \cdot y^2 \cdot \Delta x + 2 \cdot d \cdot y \cdot \Delta x \cdot \Delta y + d \cdot \Delta x \cdot (\Delta y)^2 \neq \\ \frac{\Delta f^*}{\Delta x} \cdot \Delta x + \frac{\Delta f^*}{\Delta y} \cdot \Delta y &= d \cdot y^2 \cdot \Delta x + d \cdot x \cdot (2 \cdot y + \Delta y) \cdot \Delta y \end{aligned} \quad (8)$$

the difference between the two amounting to:

$$SB = \Delta f^* - \frac{\Delta f^*}{\Delta x} \cdot \Delta x - \frac{\Delta f^*}{\Delta y} \cdot \Delta y = 2 \cdot d \cdot y \cdot \Delta x \cdot \Delta y + d \cdot \Delta x \cdot (\Delta y)^2 \quad (9)$$

Only detailed modelling/calculation can relate the functioning of  $x$  and  $y$  independently to their functioning together.

In the above,  $x$  and  $y$  were treated as independent molecules. Often they are not independent. And a change in  $x$  may also cause  $y$  to change. Then already for the linear dependence there is a complication:

$$\frac{\Delta f}{\Delta x} = a + b \cdot \frac{\Delta y}{\Delta x} \quad (10)$$

*i.e.* the functioning of  $x$  is not only given by  $a$ , but also by how  $x$  affects function through  $y$  and by  $b$ .

The paradigm of molecular biology is to determine  $a$  and  $b$ . It sees these as universal constants.  $a$  and  $b$  are the clothes of the emperor. Using the first equation, it then predicts how the molecules determine the functioning of the system. (8 shows that the prediction will be wrong whenever function depends nonlinearly on the molecular behaviour. New clothes for the emperor may not even help: the emperor should engage in a whole new game of redressing himself depending on the active and dynamic conditions he is in).

Again, where are we? Molecular biology determines  $x$  and  $y$  of the above equations, and perhaps the extents to which they change. Physiology can determine  $f^*$  and its changes. The urge is to understand why  $f^*$  changes when  $x$  and  $y$  change. The paradigm that the functional behaviour of the system can be understood from the changes in  $x$  and  $y$  by just determining  $a$  and  $b$ , is only true in linear systems. Therefore, the issue now is whether biological systems are linear, or more precisely, whether important functional properties of biological systems are linearly related to the properties of their molecules. And the issue is whether the molecular properties are independent of one another.

### 2.3 Nonlinearities and dependencies prevail in real life

There was a time where linear relations were quite popular. Substrate concentrations were assumed to be far below the  $K_M$ 's of their enzymes such that enzyme kinetics *in vivo* should be linear. Looking at the 'live' database of a number of

metabolic pathways (*cf.* [www.siliconcell.net](http://www.siliconcell.net)) one readily notes that this assumption is not realistic for the pathways that have been studied to sufficient detail experimentally. Often the substrate concentration is around the  $K_M$ , so that rates depend non-linearly on concentrations. And, in many cases kinetics is cooperative in terms of substrate concentrations, providing another reason for nonlinearity. And, almost per definition product inhibition, which occurs frequently, is nonlinear.

Indeed, Michaelis-Menten, Monod, and Hill's equations for enzyme and growth kinetics respectively are less than first-order in their dependence on concentration. Quite a few reactions depend on the concentration of more than one compound, *e.g.* when ATP is co-substrate in kinase reactions. Then the rate is bilinear or sub-linear, *i.e.* does not fulfil the linear superposition of two linear equations. Also approximations of kinetic relations in cell function reckon with nonlinear dependencies being the rule rather than the exception. Biochemical Systems Theory (Savageau 1976) uses power laws for this description with the clear intention that the powers need not equal 1. Mosaic non-equilibrium thermodynamics (MNET; Westerhoff and Van Dam 1987) uses linear relationships between the logarithms of concentrations and reaction rates, which translate into nonlinear dependencies of rate on concentrations (*cf.* Wu *et al.* 2004).

How does function depend on gene dosage? For non-redundant essential genes it is clear that function disappears upon a complete knockout. Most such genes (or rather the mutations therein) are recessive, however, this implies that half the gene dosage bestows the organism with full rather than half function: function tends to vary hyperbolically not linearly with gene dosage (Kacser and Burns 1973). The related issue of how pathway flux depends on enzyme activity can be addressed relatively strictly by Metabolic Control Analysis (MCA) and leads to a rather similar answer: pathway flux varies much less than proportionally with enzyme concentration.

There are a number of cases where it is quite obvious that simply adding the behaviour of molecules in isolation will not reproduce their behaviour *in vivo*. One is that the molecules that are involved in the cell-cycle oscillation and the molecules that are involved in glycolytic oscillations would not themselves oscillate in isolation. The oscillation only arises when many molecules of the network are present (see chapter by Novak *et al.*). Yeast glycolysis would not operate as steadily as observed experimentally if the TPS1 regulatory feedback with trehalose phosphate inhibiting hexokinase would not be active (Teusink *et al.* 1998). More generally, reaction rates through metabolic pathways can only attain steady state when they become equal to one another, which they can only do if the enzymes interact with each other, mostly by sensing the concentrations of the metabolites between them. ATP synthesis by the  $H^+$ -ATPase is only possible when the enzyme occurs in a system with electron-transfer chain linked proton pumps that generate an electrochemical potential difference for protons that is sufficiently high. If the latter are absent, ATP hydrolysis will occur. The difference between ATP synthesis and ATP hydrolysis is crucial for life.

Clearly, the more realistic assumption should be that relationships are nonlinear. Rather than assuming linearity of the dependence of reaction rates on the concentrations of their substrates and products, we use the paradigm that kinetics *in*

*in vivo* is massively nonlinear in terms of metabolite concentrations. And, we reckon that just adding knowledge of the individual molecules outside the context of the systems in which they act, will not lead us to understand their functioning in the living organism: outside the system the  $y$ 's are different, making the prediction of how  $x$  affects function, fail. This compromises 'molecular systems biology' and suggests that something is needed in addition to the massive determination of the properties of all the molecules of the living cell, individually.

### **3 Systems biology: Neither the biology of systems nor the biology of all molecules individually**

We are able to analyze function of living organisms by physiology without too much reference to molecules. We are also able to determine the properties and concentrations of most molecules that are active in living cells. However, because of the aspect of nonlinearity addressed above, adding all the bits of knowledge about the individual molecules will not lead us to understand the functioning of the living cell. And, on the other hand physiology only understands life in terms of more descriptive properties that are not clearly related to molecular events.

What is needed is a science, which we shall call 'systems biology' that connects physiology to genome-wide molecular biology, *i.e.* a science that addresses specifically how and why the system of macromolecules differs functionally from the sum of the individual behaviours of the molecules that constitute the system. Because of the importance of nonlinearities for function, this science is complex, whether we like it or not. Where physics has profited much from simplification often down to linear first-order approximations, systems biology is inherently more complicated as mostly it cannot engage in such linearization, for fear of approximating away its very essence, *i.e.* the properties that arise precisely due to the nonlinearity of the relations.

The tendency of Non-Equilibrium Thermodynamics to focus on areas where flux-force relations are quasi-linear may seem to make it more appropriate from the physics perspective than from the perspective of systems biology, were it not for the fact that a linear description in a lin-log (Westerhoff and Van Dam 1987; Wu *et al.* 2004) or a log-log world corresponds to a nonlinear world in the true coordinates, and that MNET did away with Onsager's reciprocity relations. Indeed, linear flow-force relations of non-equilibrium thermodynamics are perfectly consistent with the occurrence of oscillations (Cortassa *et al.* 1991).

The feature we noted above, *i.e.* that the dependence of function on a molecular property may depend on the intensity of that property itself and on the intensity of other molecular properties, has the implication that systems biology properties may be much more a conglomerate of special properties that are only valid in a subset of conditions than general properties. In a linear world, many completely general, condition-independent properties should be expected to dominate, but not so in a nonlinear world, although there may still be some such general properties.



And then, one should not forget that systems biology refers to ‘biology’, *i.e.* to functioning of living systems. This implies that it may not deal with the most general of all nonlinear systems, but mostly with the subset of systems that are found in biology. Accordingly, there is an emphasis on systems that are robust with respect to chemical and even evolutionary fluctuations (*cf.* Carlson and Doyle 2002; Westerhoff and Van Dam 1987). Explosive systems with unbounded reaction rates are unlikely to be common in the living cell.

Is systems biology new then? Well, *No* and *Yes* at the same time. Scientists have long studied cases where new properties arose in molecular interactions (Westerhoff and Palsson 2004). Below, we will discuss show this systems biology *avant la lettre* reached beyond both molecular biology and physiology. We will show how systems biology combines both these disciplines with at least three others, and perhaps even more. We will illustrate how systems biology has already solved issues about biology that many physiologists and molecular biologists were not even able to recognize as issues because of the limitations of their paradigms. And, we will suggest some ways in which parts of systems biology may be developed further.

## 4 Systems biology *avant la lettre*

### 4.1 Self-organization

The molecular biology paradigm sees the cell as a bag of structures kept together by a plasma membrane. The biochemistry paradigm adds that many of the structures correspond to proteins that catalyze or regulate chemical processes. However, macroscopically, and even microscopically, most living organisms do not quite look like amorphous bags of enzymes; rather, they are well structured, the top differing greatly from the bottom and perhaps the left being a mirror image rather than a replica of the right. How do these structures form? Early development of organisms is a case in point, where elaborate spatial structure arises from an apparently spherically symmetrical egg. The apparent breaking of symmetry also occurs in the dimension of time. In a continuous environment, heart cells begin to beat, cell cycles begin to run. The issue connotes with an issue of energetics, *i.e.* how chemical free energy could be converted into and from spatial free energy such as in muscle contraction, in chemicomotion of macromolecules, and in transmembrane electric potentials.

Although preceded by many others, the Brussels school led by Prigogine attracted much attention when addressing these issues (Nicolis and Prigogine 1977). It turned out that small fluctuations, which could themselves break symmetry could be amplified in some cases and lead to a less symmetric final state (both in space and in time) than the initial state. Because in particle physics there should be time invariance and conservation of momentum at the level of individual particles, this still produced a paradox. However, part of this paradox had been resolved before by statistical thermodynamics showing that in systems of many particles

some system configurations recognized as single macroscopic states actually corresponded to many more underlying microscopic states than other such macroscopic configurations. Therefore, the former are more likely to be observed. If one were to observe an unlikely macroscopic state first (*e.g.* because it had been prepared by earlier processes), then it should be highly likely that one would see that macroscopic state change to the more likely one. This is what we have learnt to call spontaneous processes. We note that this phenomenon is a property of systems of particles much more than of the particles themselves.

For these transitions between macroscopic states to occur, it is important that the system of particles can move between its microscopic configurations, to then act as if it were searching for the most probable macroscopic state. The movements between the microscopic configurations are called 'fluctuations'. They derive from the continuous bombardment of particles from the environment with energies of all sorts, at least at temperatures above zero Kelvin. The fluctuations cause a progression of a system through its many configurations, which is a random walk in terms of the microscopic states, but a much biased random walk in terms of macroscopic states. In macroscopic terms, it seems that the system exhibits a preference for the most probable macroscopic state, *i.e.* for the macroscopic state that has most microscopic configurations. The latter macroscopic state then is the final 'steady' state and such a state is then stable with respect to fluctuations, *i.e.* after fluctuations the system will return to that macroscopic state. It is this stability against fluctuations (and even of processes themselves) that is at the basis of many systems biology properties, such as robustness and the flux and control properties of dynamic systems including biological ones (*cf.* Westerhoff and Van Dam 1987). For now it is important that this very stability is a property of systems not of their components.

The above statistical thermodynamic argument is often formulated in terms of spontaneous processes producing entropy, where entropy is associated with less ordered states, or chaos. Living organisms are different in this respect as they typically produce order out of chaos, or at least maintain order. Life is, therefore, at the basis of quite a significant extension of thermodynamics where it was made clear that Gibbs (Katchalsky and Curran 1967), or rather metabolic (Westerhoff and Van Dam 1987) free energy has to be destroyed to keep life processes going, whilst some of the input free energy is transduced to free energy in the new biomass. Because this is essentially about life and biology and because it critically depends on the system nature of biology, this has perhaps been the first type of system(s) biology.

Living systems, therefore, need to operate away from equilibrium, yet not so far away that they expend all input free energy and fail to retain some of it for building their own structures. For true symmetry breaking it was shown that systems need to be further away from equilibrium than where the first order Onsager approximation applies (Nicolis and Prigogine 1977; Cortassa *et al.* 1991). Moreover, for such phenomena to occur, more than two components, asymmetrical thermokinetics, and nonlinear kinetics are needed. Such symmetry breaking or 'self-organization' can only happen in systems of molecules not in individual molecules.

## 4.2 Perpetuation

Symmetry breaking of the above, far-from-equilibrium thermodynamic type has the disadvantage that it is non-robust: it depends on the nature of the first fluctuation. If the bifurcation is to lead from spherical to left-handed symmetry for instance, one should expect an equal probability to have an outcome of right-handed symmetry. At a fifty percent error rate, this symmetry breaking would be a highly unreliable mechanism. Because many such symmetry breaking steps need to be made, this should result in little fitness *vis-à-vis* biological evolution. Indeed, experimental results now suggest that there is much less absolute symmetry breaking in biology than was once assumed.

First, most eggs are not quite symmetrical but exist in the context of asymmetrical environment set up by the maternal organism. Probably self-organization mechanisms serve to consolidate symmetry breaking and developmental decisions that have been set in motion by robust asymmetries. The latter are set in place by pre-existing biological matter. Indeed, *perpetuation* is a major characteristic of life as we know it. New cells are not created *de novo* somewhere in the middle of the maternal cell, then to be excreted as a newborn cell. Rather, the mother cell grows in volume and surface area and then either splits into two equal parts, or a small part of the mother cell pinches off and becomes the daughter cell. New proteins are synthesized on old ribosomes, new mRNA is synthesized by old RNA polymerase, and half the DNA of cells already existed in the mother cell. This perpetuation aspect of life makes the issues of symmetry breaking and self-organization much less acute. There is far less self-organization than anticipated earlier on; much of what happens is perpetuation and then division. Self-organization processes serve to maintain and stabilize decisions made through perpetuation.

## 4.3 Chemiosmotic coupling

The maternal organism may convey its own asymmetry to its daughter cell in this mass action way. However, asymmetry can also be conveyed catalytically. Proteins are asymmetrical and are inserted asymmetrically into membranes. Accordingly a cytochrome oxidase catalyzing the macroscopically scalar reaction of oxygen reduction by cytochrome c can couple this reaction to vectorial proton movement by virtue of its asymmetrical 3-D organization. A substantial fraction of the free energy in food is transduced to free energy in new biomass through the electrochemical potential difference for protons across the mitochondrial inner membrane. This involves a transition from chemical free energy, which is a scalar property, through a vectorial property back to a scalar property. Hence, it involves two changes of symmetry, the former of which involves cytochrome c oxidase. For the present chapter, it is important that of necessity, this process requires a system, *i.e.* cannot be carried out by a single macromolecule: it requires a primary proton pump generating the electrochemical potential difference of protons, a closed membrane and a secondary proton pump converting the protonmotive free

energy into ATP hydrolytic free energy. Indeed, the chemiosmotic coupling concept by Peter Mitchell (1961) was a second important case of systems biology *avant la lettre*.

#### **4.4 Non-equilibrium thermodynamics**

Many biological free-energy transduction processes have been described in terms of non-equilibrium thermodynamics. Where thermodynamics was once thought to be devoid of mechanistic detail, and indeed this was proclaimed to be an asset, it was shown that the very essence of processes that are significantly removed from equilibrium is that their thermodynamics does depend on mechanisms (Keizer 1987). MNET was developed precisely to relate systems performance to the mechanism of free-energy transduction, including those leading to incomplete coupling (Westerhoff and Van Dam 1987).

#### **4.5 Systems biology *avant la lettre*: Metabolic Control Analysis; laws of systems biology**

Non-equilibrium thermodynamics approximates rate equations by linear functions of logarithms of concentrations. This enables analytical solutions at steady state (Katchalsky and Curran 1967). In this approach, however, the parameters are treated as phenomenological and little emphasis is placed on their relationship with parameters that reflect details of mechanisms. Hill's (1977) analysis of complex biological reactions catalyzed by single enzymes showed that although their rate equations obeyed some general format, their parameter values and form also reflected mechanism. For systems of reactions, MNET (Westerhoff and Van Dam 1987) elaborated this and showed how quantitative analysis of biological free-energy transduction systems could lead to conclusions about the functioning of mechanisms such as imperfect coupling and back-pressure.

Biochemical Systems Theory (BST) approximates rate equations with power laws. For linear reaction chains, this again enables analytical solutions of the system equations for steady state (Savageau 1976). There is no emphasis on how the powers relate to underlying mechanisms; the emphasis is on qualitative systems behaviour, *i.e.* on physiology in the above definition. In addition, BST and MNET result in descriptions and sometimes tendencies for systems to behave in certain ways, but not in general principles or 'laws'.

Metabolic Control Analysis (MCA) is both less and more ambitious than BST and MNET. First it does not aim at describing the entire dependence of functional properties on process properties. It focuses on the infinitesimal dependence of functions on those properties. The disadvantage of this is that MCA only considers small changes in the system. The advantage is that in this, MCA is not an approximation but exact, and that as a corollary thereof, MCA has derived some general principles for metabolic systems, *i.e.* some systems biology 'laws'. We shall here illustrate one of these, *i.e.* the summation law for the control of flux

through a metabolic pathway by the various reaction steps in that pathway. As in biology almost all reactions are catalyzed by enzymes, the control by a reaction step is related to the control by proteins and to the control by gene expression. The control of the steady-state flux  $J$  through a metabolic pathway, such as in Fig. 1, by enzyme  $i$  is quantified in terms of the so-called control coefficient of that enzyme *vis-à-vis* that flux:

$$C_{e_i}^J \equiv \left( \frac{d \log |J|}{d \log e_i} \right)_{e_j, j=1, \dots, n, j \neq i} \quad (11)$$

where  $\log$  stands for logarithm with any base.  $e_i$  refers to the catalytic activity of the  $i^{\text{th}}$  step in the pathway and can in simple cases be replaced by either the  $V_{\max}$  or the concentration of the enzyme catalyzing that reaction. Technically speaking, the  $d$  here stands for a partial derivative, with the conditions that the other process activities are held constant and the steady state is re-attained. The logarithmic derivative is taken at the physiological state and can be replaced by the normal derivative provided that the result is then multiplied by the ratio of enzyme activity to flux.

The control coefficient quantifies the importance of the  $i^{\text{th}}$  step in the pathway for the pathway flux. Taking the molecular biology point of view to the extreme one might wish to determine this flux control coefficient *in vitro* for every enzyme individually and then assume that that control coefficient should be approximately the same *in vivo*. With the enzyme in isolation, the flux through the enzyme is directly proportional to its concentration, which implies that the flux control coefficient equal 1. For a pathway of  $n$  enzymes, this should imply that all enzymes should be flux-limiting and that the sum of all flux control coefficients should equal  $n$ . MCA falsifies this conjecture: it has a law that says that the sum of the flux control coefficients for any flux over all processes equal 1:

$$\sum_{i=1}^n C_{e_i}^J = 1 \quad (12)$$

Contemplating a two-step metabolic pathway, it is simple to see how this relates to the issue that systems biology deals with nonlinear systems in which processes are dependent:

$$\frac{dJ}{de_2} = \left( 1 - \frac{e_1}{J} \cdot \frac{dJ}{de_1} \right) \cdot \frac{J}{e_2} \quad (13)$$

*i.e.* the functional property flux ( $J$ ) neither depends on the molecular properties  $e_1$  and  $e_2$  independently, nor on either of them linearly. The more intuitive explanation for this is that when one increases the activity of one enzyme in the pathway to see if it is flux limiting, one simultaneously makes the other enzymes more flux limiting.

Laws of MCA address the sum totals of control coefficients with respect to flux and concentration, as well as relationships between control coefficients and enzyme properties called elasticity coefficients. The latter laws are called connectivity theorems and relate to the stability of the system against fluctuations (*cf.* above and Westerhoff and Van Dam 1987). They are often called theorems but

would equally qualify as ‘laws’ of systems biology, as they indeed address the difference between properties of the systems and properties of the molecules in isolation.

The enzyme properties that are important in the connectivity laws are the elasticity coefficients. They denote the extent to which a reaction step, hence the enzyme that catalyzes it, responds to changes in its metabolic environment. For the elasticity of enzyme 1 of Fig. 1 with respect to the concentration ( $z_2$ ) of metabolite  $Z_2$ :

$$\varepsilon_{z_2}^1 \equiv \frac{\partial \log |v_1|}{\partial \log z_2} \quad (14)$$

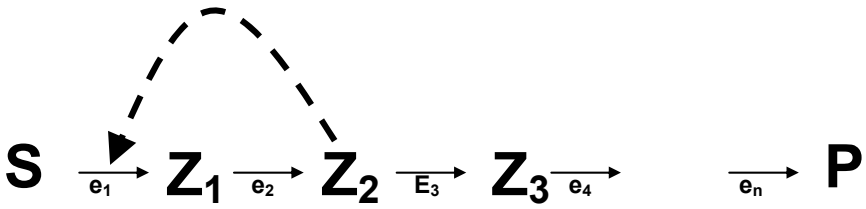
For the case that enzymes 1 and 2 of the pathway are product insensitive, but enzyme 1 is sensitive to the concentration of metabolite  $Z_2$ , one finds for the control of the steady state flux through enzyme 1:

$$C_1^J = \frac{\varepsilon_{z_2}^3}{\varepsilon_{z_2}^3 - \varepsilon_{z_2}^1} \quad (15)$$

When enzyme 1 is well regulated by metabolite  $Z_2$ , the corresponding elasticity  $\varepsilon_{z_2}^1$  is quite high in absolute magnitude. Through the above equations, this has the effect that enzyme 1 exerts no control on its own steady state reaction rate; all that control may then reside in enzyme 3 of figure 1. In this way, MCA can illustrate that the *in vitro* control a macromolecule has of its own activity, may be absent in the physiological situation of the intact system. The other factors, residing in different macromolecules, may control its function. This brings home the key issue of systems biology that important aspects of cell function reside in the interaction properties (the elasticity coefficients) rather than in the properties of the individual molecules. It also reinforces the role of MCA as an important theoretical tool in systems biology.

#### 4.6 Circular causality and emergence

Biochemistry and molecular biology reinforced the use of the scientific methods of contemporary physics and chemistry in biology, methods which arose from and are still closely tied to the Newtonian view of the world. The Newtonian perspective can be reconciled with three of the four Aristotelian causes, *i.e.*, the state of the system at time  $t+\Delta t$  can be explained as the effect of a material cause corresponding to the state of the system at time  $t$ , an efficient cause corresponding to the mathematical form of the recursive state transition function, and a formal cause corresponding to the initial state at time  $t_0$  and other parameter-values. However, as argued by Rosen (see *e.g.* Rosen 1991), there is no place in this view for the fourth Aristotelian causal category, namely that of final cause. The reason



**Fig. 1.** Metabolic pathway, with feedback from the second metabolite on the first reaction. We shall assume all reaction irreversible and not product inhibited.

for this is that whereas material, efficient and formal causes all work in the forward direction, final cause works backwards, which in the Newtonian framework would imply the future affecting the present. From the perspective of physics, chemistry and molecular biology, final cause is, therefore, illegitimate.

Final explanations are closely linked to the concept of function, which is indispensable for many explanations in mainstream biology, and is, despite its uncertain status in formal arguments, often invoked as an inspiration for finding phenomena and even mechanisms. Today many biologists regard evolution and selection of the fittest as an acceptable basis for the use of final causation as part of a scientific explanation. The more frequent type of argument is that a certain mechanism is in place *because* it leads to a higher growth rate or to a higher growth efficiency. The background for this argumentation is often left implicit but is taken to be that because the mechanism improves growth rate, it should have been selected for in evolution, which explains why it is present in the organism under study: ultimately the explanation is then effectively rephrased in terms of formal causation. Common examples of application of final cause include statements such as the occurrence of multidrug resistance proteins at the blood-brain barrier, because it helps to keep toxins out of the brain. Recent examples in systems biology include the flux balance analysis of Palsson and colleagues (Reed *et al.* 2003). Here fluxes are calculated on the basis of earlier experimental results on metabolic pathway genes specifying the metabolic network, and on the assumption of maximal growth rate (which thereby acts as a final cause). The correspondence of the calculated fluxes with experimental fluxes is taken to indicate the correctness of the pathway model.

The first three Aristotelian causes appear to be more solid than final cause. On the other hand, one can see particularly in biology that the implementation of final causes in scientific research can be useful, provided that the final cause (such as the assumed requirement to be optimal in terms of growth rate) is mentioned explicitly. If then, later, it turns out that an organism has not been optimized for growth rate, the argumentation drops for that particular organism but may remain in place for others. After all, other parts of science, with mathematics as a champion, operate on the basis of axioms, which are assumed material causes, and it is accepted that such sciences are useful only for phenomena that obey those axioms.

However, biology may not yet be quite ripe for the acceptance of final cause. First it needs to be certified for every final cause that it has indeed been brought

into play in evolutionary selection. Although, it may seem that evolution is able to select for any odd property that could in theory enhance survival potential, this is not so. For instance, lions born with jet engines would be much more successful in hunting, but such lions could certainly not emerge through biological evolution. An understanding of whether natural selection could have indeed selected for a presumed function again requires insight in the functioning of the entire biological system: it requires systems biology. For as long as insufficient systems biology is in place, it is perhaps still better to weed out the implicit use of final (forward) causation that is all too frequent in a biology that pretends to follow rationality only.

The above sketches the way mainstream biology currently attempts to cope with final causation. Contemporary physics and chemistry, on the other hand, are quite strict in eradicating final causes in their scientific arguments. Even more strongly they consider as illegitimate the circular causation implied in explanations in which we say that *A* causes *B*, *B* causes *C* and *C* causes *A*. However, Rosen (1991) argued persuasively that it is exactly this type of causation that distinguishes living organisms from non-living systems. In brief, Rosen demonstrated that for living systems to be able to fabricate themselves autonomously (*i.e.* be autopoietic in the sense of Maturana and Varela 1980) they need to be organised in a way such that all efficient causes are inside the system. Using the formalism of category theory, he developed a *relational biology* with which this living organisation can be described.

There are other types of circular explanation that have more to do with causation in the Humean sense of one event causing another subsequent event; these also have potential value in biology. Examples are found in induction of gene expression and in the cell cycle: lactose uptake in *E. coli* causes an increase in intracellular allolactose, which binds lac repressor and causes induction of the lac operon which causes enhanced expression of lactose permease, which causes enhanced uptake of lactose, etc. Accordingly, lactose uptake causes (more) lactose uptake. In glycolytic oscillations, activation of phosphofructokinase causes an increase in AMP which causes a further such increase. The ensuing stronger drop in ATP and increase in fructose bisphosphate causes the lower part of glycolysis to make more ATP, which then again stimulates phosphofructokinase which then again decreases ATP and increases fructose bisphosphate which stimulates the lower part of glycolysis. Here phosphofructokinase activation causes phosphofructokinase activation, be it with a time delay; oscillations being the consequence.

Because life is a self-sustaining phenomenon, it has mechanisms in place that cause effects that in turn cause their causes. Although this may not fall within the accepted paradigm of the physical chemical sciences, the phenomenon of circular causation may be so essential to biology that it should be dealt with. As mentioned above, Rosen (1991) has shown how to deal with organisation that leads to circular fabrication. Below we shall show a way that deals with circular event-causality by dissecting it into two or more parts, employing mathematical methodologies.



## 4.7 Networks and hierarchies in life

The identification of most genes encoding the metabolic enzymes of some organism has enabled methodologies for the systematic mapping of metabolic networks. The method first identifies the genes that encode enzymes, then identifies the chemical reactions these enzymes catalyze, and then writes for each enzyme which chemical compounds it produces and consumes, and at which stoichiometry. The stoichiometries for each reaction are then denoted as the column of a huge stoichiometry matrix  $N$ . The multiplication of  $N$  with a vector  $v$  of all reaction rates then gives the time dependence of the concentrations of all chemical substances ( $dm/dt$ ) inside the cell:

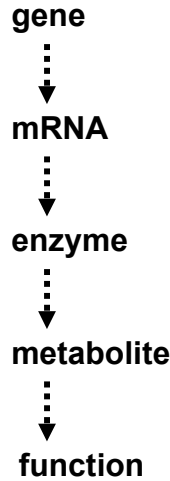
$$\frac{dm}{dt} = N \cdot v \quad (16)$$

One then tries to determine which combinations of reaction rates should make the right-hand side of the equation equal zero; those are the rates that are compatible with steady states. Technically, these rates are in the Kernel of matrix  $N$ , and they may well be ones that are biochemically unrealistic for instance by proceeding thermodynamically uphill. In addition, the number of possible reaction rates that lead to steady state is very large. Schuster and colleagues added the requirement that all reactions proceed as allowed by thermodynamics, leading to the so-called elementary reaction modes of the network (Schuster *et al.* 2000). By also taking into account the stoichiometries at which external substrates are utilized and external products are formed, this enabled Schuster and colleagues to examine whether the known network was able to produce certain chemical compounds from certain substrates. Palsson and colleagues added other restrictions such as maximum capacities and later (see above) maximum efficiency or rate of growth (Reed *et al.* 2003). This led them to unique solutions for  $v$ , which were often close to experimental observations.

These pieces of work are examples of systems biology, in the sense that they depend completely on the connectivity of the network; for an individual reaction such an analysis is impossible. In addition and in contrast to our earlier examples of systems biology, they rely on most, if not all, of the network being known, *i.e.* they are genome-wide.

Both these methods focus on network topology, neither takes kinetics into account. What they, therefore, obtain is *potential* flux patterns, *i.e.* flux patterns that would materialize if indeed kinetics of all the individual reactions were such that metabolite concentrations could evolve to steady state values that are consistent with all those fluxes.

The ‘central dogma’ of molecular biology states that DNA makes mRNA makes protein makes metabolites. Literally this statement is illustrated in figure 2. Kahn and Westerhoff (1991) observed that what was meant was actually orthogonal to the literal interpretation of the statement: DNA does not make mRNA; rather, RNA-polymerase (efficient cause) makes mRNA (effect) from nucleotide triphosphates (material cause) using DNA as template (formal cause). Likewise,



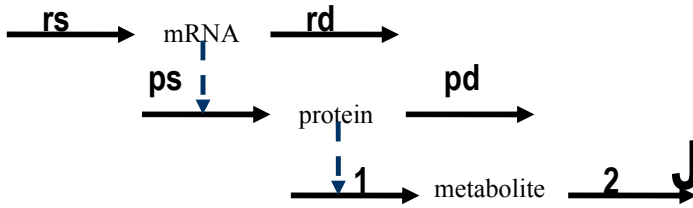
**Fig. 2.** The central dogma of molecular biology: DNA gives mRNA gives protein, gives catalytic activity, gives metabolites and function. 30,000 such diagrams would represent molecular systems biology in humans.

ribosomes, rather than mRNA, make proteins from amino acids using mRNAs as templates. These authors then emphasized that in essence cell function is subdivided in various hierarchical levels, one at the level of mRNA metabolism, one at the level of protein metabolism and one at the level of intermediary metabolism (*cf.* Fig. 3). In essence (though not quite strictly), these levels are *not* converted into each other but regulate conversions inside each other. This means that the stoichiometry matrix  $N$  (*cf.* above) is block-diagonal. This feature enabled the development of a new version of MCA called Hierarchical Control Analysis, which led to new laws specific for such hierarchical systems. One of these laws is that:

$$C_{rs}^J + C_{rd}^J = 0 \quad (17)$$

acknowledging that transcription of the gene encoding the protein that catalyzes reaction 1 can also control the flux  $J$ , and that (at steady state) this control is equally strong as the control by the process that degrades this mRNA. The strength of these controls can readily be 1 and -1 respectively, leading to the effect that the flux through step 2 of the pathway of Fig. 3 can be strongly controlled by a process that is quite remote in the cell's network, *i.e.* the degradation of the mRNA of enzyme 1. This again illustrates that the processes run by macromolecules (such as the enzyme catalyzing step 2 in this example) in living systems, are usually not determined by these macromolecules themselves, but rather by the interactions of all the macromolecules, *i.e.* by what makes the system differ from its components.

Fig. 3 shows how metabolism is not only determined by metabolic control but potentially also by transcription and translation control. It suggests a hierarchy of control, transcription presiding over translation, which would again preside over



**Fig. 3.** Hierarchical organization of cell function, as simplified for the function flux ( $J$ ) through a two-step metabolic pathway, leading through reaction 1 from a substrate  $S$ , at fixed concentration, not shown, through the metabolite at a variable concentration, through a reaction 2 to a product  $P$ , at constant concentration, now shown. Only of step 1 it is shown explicitly that it is catalyzed by a protein, which is synthesized in a ‘protein synthesis process ‘ps’ and degraded in a protein degradation process ‘pd’. The protein is not converted to the metabolite, hence, the dashed arrow from protein to reaction 1. The synthesis of this protein occurs in a process that is specified by the corresponding mRNA. This specification, however, corresponds to an influence not to a conversion of mRNA into the protein, hence, the dashed line. The mRNA is synthesized in a process called ‘rs’ and degraded through a process called ‘rd’. Note that feedback from the lower to the upper levels is not taken into account here; this is a dictatorial system.

metabolism. Signal transduction networks also consist of various levels of organization that are not connected by mass flow, just by information flux. Many of the laws Hierarchical Control Analysis also pertain to signal transduction and, many more are still being discovered (Hornberg *et al.* 2005).

The type of control structure in figure 3 has been called a ‘dictatorial’ control hierarchy. Such dictatorial control systems are highly robust against internal fluctuations, but not adaptive if anything goes wrong structurally. Most biological systems appear to be more sophisticated in that the upper level in the hierarchy (*i.e.* transcription) is not autonomous but adjusts itself to altered requirements at the metabolic, *i.e.* functional level: Transcription regulation often responds to changes at the metabolic level. For instance, allolactose in *E. coli* induces the enzyme that metabolizes it (indirectly, see chapter by Kremling *et al.*) with the effect that if the cell sees lactose (and is able to take it up, see above) it will synthesize more of the enzymes involved in the corresponding catabolic pathway, but only if insufficient such enzymes are present, *i.e.* only if allolactose accumulates to some extent. Biologists have become used to this phenomenon, but perhaps not to the implications that (i) it compromises the central dogma of molecular biology somewhat (*cf.* Fig. 2) in that now metabolism also determines gene expression, and (ii) it provides biology with circular causation, metabolism being the cause of changes in gene expression which in turn causes changes in metabolism.

## 4.8 Systems biology: dealing with the circular causation in biology

Systems with circular causality are shied way from by physics, chemistry, and molecular biology and for good reasons: such systems are much more difficult to handle experimentally and their analysis runs the risk of either becoming futile when the feature of circular regulation is neglected, or becoming inconsistent because cause and effect are interchanged inappropriately.

Again, by taking the feature of circular regulation into account explicitly, systems biology has to deal with one of the features that perhaps distinguish biology most from physics and chemistry. We shall now show how Hierarchical Control Analysis as one of the systems biology approaches *avant la lettre*, is able to deal with the issue of circular regulation. Basically, it does this by cutting away a regulatory link that causes the regulation to be cyclic, by then analyzing the regulation of the remaining non-cyclic parts independently, and by then using mathematics to understand the circularity that occurs in the intact system.

We shall illustrate this for a simplified version of Fig. 3, *i.e.* Fig. 4, which still contains the circular regulation but has neglected the level of translation (protein synthesis and degradation). We shall consider the extent to which enzyme 1 controls its own flux, *i.e.* to what extent the flux through enzyme 1 changes when we activate the  $V_{max}$  of that enzyme (through a point mutation, or by reducing the concentration of a non-competitive inhibitor). This increase in its  $V_{max}$  will cause an increase in X, which may activate the transcription of gene 1 which will then lead to an increased amount of enzyme 1; activation of step 1 will have been caused by activation of step 1 (we, here, consider the feedback loop to be positive; when the feedback is negative, circular causation is negative, less obvious, but perhaps even more confusing).

We dissect the overall system into two subsystems, one at the level of mRNA metabolism and one at the level of the metabolic pathway. We do this by setting the effect of X on transcription to zero, *i.e.*, by assuming that

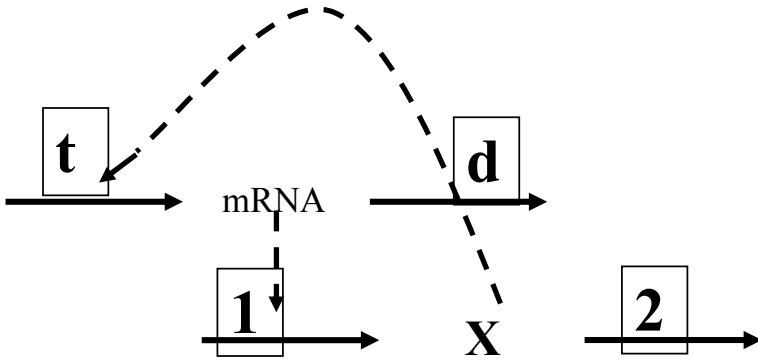
$\varepsilon_{X}^t \equiv \frac{\partial \log |v_{translation}|}{\partial \log [X]} = 0$ . The control coefficients obtained for this dissected

system are written as lower-case c's. For the control exerted by enzyme 1 on the metabolic flux this amounts to:

$$c_1^J = \frac{\varepsilon_x^2}{\varepsilon_x^2 - \varepsilon_x^1} \quad (18)$$

For the dissected control of enzyme 1 on the concentration of metabolite X, one finds:

$$c_1^x = \frac{1}{\varepsilon_x^2 - \varepsilon_x^1} \quad (19)$$



**Fig. 4.** Simplified hierarchical organization of cell function. The transcription and translation levels have been contracted to a single level overlying the metabolic level, in which enzyme 1 synthesizes metabolite X and enzyme 2 degrades it. At the upper level mRNA is synthesized by process ‘t’ and degraded by process ‘d’. The dashed arrow from mRNA to enzyme 1 refers to the transcription-translation regulation of the level of enzyme 1. The dashed level from X to ‘t’ refers to transcription regulation by the level of metabolite X. The latter regulation makes the network ‘democratic’.

For the control exerted by transcription on the concentration of mRNA for enzyme 1, it leads to:

$$c_t^{mRNA_1} = \frac{1}{\varepsilon_{mRNA_1}^d - \varepsilon_{mRNA_1}^t} \quad (20)$$

For the intact system, Hierarchical Control Analysis has shown that for the control exerted by the  $V_{max}$  of enzyme 1 on the flux one finds (Kahn and Westerhoff 1991; Hofmeyr and Westerhoff 2001):

$$C_{V_1}^J = \left( \frac{d \ln J}{d \ln V_1} \right) = \frac{c_1^J}{1 - \rho} = \frac{c_1^J}{1 - c_t^{mRNA_1} \cdot \varepsilon_X^t \cdot c_1^X} = \frac{c_1^J}{1 - \frac{1}{\varepsilon_{mRNA_1}^d - \varepsilon_{mRNA_1}^t} \cdot \varepsilon_X^t \cdot \frac{1}{\varepsilon_x^2 - \varepsilon_x^1}} \quad (21)$$

The term:

$$\rho = c_t^{mRNA_1} \cdot \varepsilon_X^t \cdot c_1^X = \frac{1}{\varepsilon_{mRNA_1}^d - \varepsilon_{mRNA_1}^t} \cdot \varepsilon_X^t \cdot \frac{1}{\varepsilon_x^2 - \varepsilon_x^1} \quad (22)$$

is the circular causation term. It quantifies the regulation of enzyme 1 (through  $mRNA_1$ ) by transcription, multiplied by the regulation of transcription by X, multiplied by the regulation of X by enzyme 1. If the circular causation term is  $-1$ ) this halves the control the  $V_{max}$  of enzyme 1 exerts on the flux, corresponding to a case of homeostatic regulation. If the circular causation term is plus 1, or higher, then the circular causation causes instability and perhaps self-organization, this is a bifurcation point (*cf.* chapter by Novak *et al.*).

This example shows how Hierarchical Control Analysis can deal with circular causation. An important aspect is that circular causation can have one out of various possibly strengths. Only for some such strengths, it may become difficult to analyze the system. But for most others, circular causation can be analyzed quantitatively and is seen to adjust the robustness, *i.e.*, homeostasis of the system.

## 5 Concluding remarks

What is systems biology then? And if there was systems biology *avant la lettre*, what new is there under the systems biology sun?

Systems biology studies, in a fully scientific manner, the functional properties that arise in the dynamic nonlinear interactions between the components of living systems. This implies that it is based both on experimentation and on strict criteria of scientific testing of theories. There should be no open ends, and the system under study should be characterized completely, if not immediately then at least ultimately. Therefore, although in the initial building stages, systems biology may limit itself to fairly autonomous parts of living cells (*cf.* [www.siliconcell.net](http://www.siliconcell.net)): this is only whilst *en route* to the analysis of the complete living cell. Systems biology is tied in strongly with genomics, proteomics, and metabolomics, in the sense of the ability of measuring all concentrations cell-wide.

Systems biology is a science in and of itself. Hence, it does not boil down to modelling of part of a living cell, or to measuring all metabolite concentrations in that cell, however important each may be. Systems biology tries to discover new principles behind the functioning of living organisms. Genome-wide experiments and models of parts of living cells are tools in that discovery, not aims in themselves.

The systems biology *avant la lettre* mentioned here has shown that important principles can indeed be discovered, but that appreciable parts of the living cell remain to be explored. In addition, the already discovered principles have not yet been examined in terms of their validity or usefulness for the genome size systems. And, inspection of the larger systems, in terms of their functionality may indeed lead to principles that do not reign in the smaller, theoretical systems studied thus far (as suggested for instance by the study of Reed *et al.* (2003).

## References

- Carlson JM, Doyle J (2002) Complexity and robustness. *Proc Natl Acad Sci USA* 99 Suppl 1:2538-2545
- Cortassa S, Aon MA, Westerhoff HV (1991) Linear non-equilibrium thermodynamics describes the dynamics of an autocatalytic system. *Biophys J* 60:794-803
- Hill TL (1977) *Free Energy Transduction in Biology*. Academic Press New York
- Hofmeyr J-HS, Westerhoff HV (2001) Building the cellular puzzle: Control in Multi-level reaction networks. *J Theor Biol* 20:261-285

- Hornberg JJ, Bruggeman FJ, Binder B, Geest CR, Bij de Vaate AJM, Lankelma J, Heinrich R, Westerhoff HV (2005) Principles behind the multifarious control of signal transduction ERK phosphorylation and kinase/phosphatase control. FEBS J 1:244-258
- Kacser H, Burns JA (1973) The Control of Flux. Symp Soc Exp Biol 27:65-104
- Kahn D, Westerhoff HV (1991) Control theory of regulatory cascades. J Theor Biol 153:255-285
- Katchalsky A, Curran P F (1967) Non-equilibrium thermodynamics in biophysics. Harvard University Press Cambridge MA, USA
- Keizer J (1987) Statistical thermodynamics of non-equilibrium processes. Springer-Verlag Berlin
- Maturana HR, Varela FJ (1980) Autopoiesis and cognition: The realisation of the living. D. Reidel Publishing Company Dordrecht
- Mitchell P (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism. Nature 191:144-148
- Nicolis G and Prigogine I (1977) Self-organization in nonequilibrium systems. Wiley and Sons, New York
- Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12. Genome Biol 13:2423-2434
- Rosen R (1991) Life itself. Columbia University Press New York
- Schuster S, Fell DA, Dandekar T (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18:326-232
- Savageau MA (1976) Biochemical systems analysis. Addison-Wesley Reading MA
- Teusink B, Walsh MC, Van Dam K, Westerhoff HV (1998) The danger of metabolic pathways with turbo design. Trends Biochem Sci 23:162-169
- Westerhoff HV, Palsson BOP (2004) The evolution of molecular biology into systems biology. Nature Biotechnol 42:1249-1252
- Westerhoff HV, Van Dam K (1987) Thermodynamics and control of biological free-energy transduction. Elsevier Amsterdam
- Wu L, Wang W, van Winden WA, van Gulik WM, Heijnen JJ (2004) A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics. Eur J Biochem 271:3348-3359

Hofmeyr, Jan-Hendrik S.

Dept. of Biochemistry, University of Stellenbosch, Private Bag X1, Matieland 7602, Stellenbosch, South Africa

Westerhoff, Hans V.

Institute for Molecular Cell Biology and Swammerdam Institute for Life Sciences, BioCentrum Amsterdam, De Boelelaan 1087, NL-1081 HV Amsterdam, EU

[hans.westerhoff@falw.vu.nl](mailto:hans.westerhoff@falw.vu.nl)