Systems Biology
Editors: Fred C. Boogerd, Frank J. Bruggeman, Jan-Hendrik S. Hofmeyr and Hans V. Westerhoff

2
The methodologies of systems biology

Hans V. Westerhoff and Douglas B. Kell

School of Chemistry, The University of Manchester, Manchester M60 IQD, UK

SUMMARY

In this book on philosophical aspects of systems biology, this chapter summarizes the philosophical status of a variety of sciences. Biology, physics and molecular biology offer particular contrast here. It is contended that philosophy and methodology should be determined substantially by the degree of complexity of the system under study. Some of the new experimental methods that have made systems biology possible are summarized. Research strategies that claim to be systems biology yet approach the topic in different ways are described. Inductive reasoning and the development and exploitation of suitable technologies are important parts of the systems biology agenda but are not themselves hypothesis-dependent science. A new methodology for systems biology is sketched that spirals in an iterative manner between experiments and theory but makes inherent use of mathematics in ways that are new to the life sciences. It is shown that the construction of a computer replica of parts of living systems has become possible and that the 'silicon cell' strategy enables the calculation of emergent properties. This may then serve as a basis for subsequent discussions with philosophers of science about how new and unique the philosophical foundations of systems biology are or should be.

1. THE METHODOLOGY AND PHILOSOPHICAL FOUNDATIONS OF THE VARIOUS SCIENCES

1.1. Physics

According to classical philosophy of science (e.g. Carnap, 1966; Nagel, 1961), science advances by an iteration between the world of mental constructs (ideas,

background knowledge, hypotheses) and the world of sense data (experimental observations). Laws (theories, hypotheses) are induced from empirical findings (Carnap, 1966). Consequences deduced by combining hypotheses with established underlying principles (such as fundamental laws of chemistry and physics) are examined experimentally to test the new hypotheses (see also Fig. 3). Given sufficient positive testing, they are transformed to underlying principles through theorization. For testing, theories should be quantitative (Carnap, 1966). It is seen as a great asset when laws and theories can also be *reduced* to underlying theories of greater validity and generality. Here thermodynamics has always served as an example; its first and second laws were first determined empirically (Nagel, 1961). The former was then elevated to a general scientific law that is also valid at the more microscopic level. The latter was deduced from the underlying principle of large numbers of substates and evolution towards increased probability with time. Quantum mechanics has also served as such an example: Schrödinger's equation and wave functions were 'induced' so as to be able to explain observations, such as the periodicity in the Table of Chemical Elements. Modern elementary particle physics appears to continue along these lines, ever inducing new phenomena and properties such as quarks, charms and colours. More generally, physics aims to explain multiple phenomena on the basis of simpler and fewer principles. Indeed, the first law of thermodynamics is much simpler than the 100% efficient conversion between all sorts of energy that it prescribes. In the classical philosophy of science, explanation by simple underlying principles is important (cf. Nagel, 1961, p. 321).

Of course, this philosophy of science is incomplete. It is very often too simplistic to induce predictions from hypotheses that can be verified. Indeed, it is seen in most quarters as much more important to try to make predictions that can then be used to falsify hypotheses (Popper, 1992). Then in practice, the sociology of science also comes in, where hypotheses are not actually falsified by their originators, but rather by competing, younger researchers, albeit only after the proponents of the original hypothesis have become less active or passed on (cf. Kuhn, 1996; Lakatos, 1978; Primas, 1981). However, this is not the issue we would like to discuss here, as we shall focus on the extent to which classical, molecular and systems biology do conform to what used to be defined as science by the main philosophers of science, or more specifically physics (Carnap, 1966).

1.2. Biology

While theoretical physics is both respectable and a major part of the activities of physicists, theoretical biology is a minor part of modern biology and is treated largely with disdain by most experimentalists (Kell, 2006). Not all of classical biology conformed strictly to the scientific methods recalled above, as it was

largely observational (Brent, 1999). Much of that science of biology accepted the diversity that appeared to inhabit the biosphere: organisms were classified and compared, and their behavior was studied in the sense of establishing correlations between properties. These correlations were rarely put to the test in the sense of falsification or even verification; observations were dominant; laws, even phenomenological ones, were rare. Classificatory concepts sufficed (Carnap, 1966).

Physicists were much stricter; they expected their codifications to produce immutable laws. Thus, the type of biology being studied caused many physicists to disdain biology, which would then be seen as an 'other science' if a science at all. Biology was 'stamp collecting', and it was claimed that physics was superior. Those who have witnessed field biologists efficiently recognizing birds in complex ecosystems, and predicting with an 80% success rate what the individual birds would do next, are perhaps less convinced of the truth of the dictum of the physicists. After all, the complexity of the prediction made by the biologist and what one might consider the total success of that prediction (i.e. success rate multiplied by complexity) was many times higher than that of the physicist predicting the probability of the location of an electron on the basis of a wavefunction. Interestingly, chemistry and biochemistry have always been middlemen; although chemistry was claimed to be a science conforming to the principles proclaimed by the philosophers of science, it often was not; organic chemistry, for instance, was rule-based rather than theory-based, albeit fairly successful in predicting possible chemical reactions and reaction mechanisms. Chemistry warrants its own philosophy of science, distinct from that of physics (Primas, 1981).

We suggest that the basic problem of biology at that time, and to some degree now, which distinguished it from the objects of study surveyed by physicists, was that the object of their study, i.e. life, was too complex to be amenable to the 'Physics' of Rutherford. The number of unobserved and in fact unobservable degrees of freedom was virtually unlimited. Every possible hypothesis would always be falsifiable, as there could always be exceptions, or additional unknown components of the system that would perturb the rule (the 'hidden variables' of certain approaches to understanding the behaviour of quantum systems). Even Mendel's 'laws' were subject to many exceptions, and it is now all too easy to scorn Mendel for overemphasizing the overall principles and for down playing the aberrations (it is also widely accepted that Mendel's data were 'too good to be true'). What would have happened with Newton discovering the laws of classical mechanics had the velocity of light been 0.1 m/s? Then Newton would have been plagued by apparent exceptions (because of relativistic corrections). Or what would have happened if all the objects around us had had substantial Coulombic charge, so as to perturb the observation of $F = ma$, in those days at least?

Classical (Organismal (Nagel, 1961)) biology was (and is) a science in that it obeyed strict methods, was devoid of unfounded predictions and aimed for reproducibility. It was, however, seen as incomplete in that its predictions were often perturbed by unexpected variations. On the contrary, it did not shy away from studying the complex and the most interesting phenomena in existence, i.e. life.

Much (though not all (Primas, 1981)) of physics did conform to the scientific methods delineated by the classical philosophers of science. How could it? Well, first of all it studied objects that happened to be simpler than the objects studied by biology; billiard balls, protons and electrons are inherently simpler than haemoglobins, monkeys and tumor cells. Certainly, it has been an extreme challenge to mankind to understand the circling of electrons around conglomerates of protons and neutrons, but the scientific achievements have been enormous. However, the number of degrees of freedom involved in the explanations of physics has been much smaller than the number of degrees of freedom in the objects of biology. Physicists (and engineers) sought this simplicity; they preferred to study single objects or systems with very few degrees of freedom, and preferably linear interactions. This enabled the discovery of simple principles and their codification by analytical mathematics. Physics could be physics and not stamp collecting, precisely because physicists selected a particular subset of stamps rather than the most beautiful and extensive stamp collections as objects of study.

This focus on simpler systems and the emphasis on simple principles, often enforced by first- and perhaps second-order linear approximations, have been very good for the development of science. Enormous progress was made for those objects of study that were simple in the above sense. Doubts arose when others noted that many problems in the environment around us were not being solved by physics. These included the weather, the behavior of the stock market, the behavior of the majority of (nonideal) gases, and life and disease.

When confronted with those issues, some physicists reversed the argumentation. It was not physics itself that was unfit to study those systems that were more complex. Rather, those objects of studies were unfit for pure physics; they might perhaps be studied by applied, less pure physics, perhaps through simulation of all the special cases. Nonequilibrium thermodynamics of the Westerhoff (Westerhoff & van Dam, 1987) type, nonequilibrium statistical mechanics of the Keizer type (Keizer, 1987) and later the discovery of deterministic chaos (e.g. Gleick, 1988) were such ‘impure’ physics. On the contrary, they demonstrated that many aspects of reality may be beyond the understanding of simpler physical theory. Prigogine was a case in point, searching for a general principle of nonequilibrium steady states in arbitrary systems, which does not exist (Nicolis & Prigogine, 1977). Some physicists moved towards biology, accepting that physics itself should change and adopt complexity. Terrell Hill is one of these,

being attracted to biology because its phenomena were inherently interesting and developing physics methods so as to be able to deal optimally with its complexity (Hill, 1977). Much of modern physics of course does accept the complexity and is subject to the limitations of nongenerality and nonlinearity plaguing biology (Fröhlich & Kremer, 1983; Primas, 1981). In this sense, we admit that we here caricature physics to serve as a contrast in a description of the essence of systems biology.

1.3. Biochemistry and molecular biology

Whilst it was welcome that physics was able to deal so elegantly with a number of phenomena, the problem for science was that much of what is inherently interesting to mankind appeared to be left intractable. Life itself, in the sense of understanding the material basis of the functioning of living organisms, therewith eluded the science that followed the methodology of physics (Rosen, 1991). There could be only two ways out of this dilemma: either physics adapted to life as an object of study, or the object of study, 'life' was adapted to the methodology of physics (perhaps with new, superphysical laws to be added, as in Schrödinger's agenda (Schrödinger, 1944, p. 80)). The latter strategy has been the basis of yet another success story, i.e. that of biochemistry, biophysics and molecular biology. It was indeed set in motion by physical scientists such as Michaelis and Menten, Franklin, Watson and Crick. Michaelis and Menten set out to study the reaction catalyzed by a single protein, while Franklin, Watson and Crick looked at a piece of a double-stranded DNA molecule. The molecular processes carried out by macromolecules in living organisms were characterized in this manner. In addition, simple and qualitative schemes of how they function together were drawn as cartoons (such cartoon-based modelling was and is a significant part of these sciences (Kell & Knowles, 2006)). This includes the one showing that a piece of DNA contains the inheritable information, which can be expressed through mRNA into proteins, which then carry out function by catalyzing metabolic conversions, signalling and work. In these three disciplines of biochemistry, biophysics and molecular biology hypotheses were proposed and verified experimentally.

However, although they tried and claimed to operate in accordance with the methodology of physics, as time proceeded, biochemistry and molecular biology became less and less anchored on the principles expounded by chemistry and physics. The hypotheses and the activities of molecular biology became intentionally largely qualitative, and the concepts comparative (Carnap, 1966), so that their tests (verifications/falsifications) could give a digital yes/no answer. With this and with a strong tendency to empirical-rather than hypothesis-driven science, biochemistry and molecular biology became immensely successful. It is now possible to purify many or most of the water-soluble macromolecules that

are active in living cells and determine their structure by X-ray crystallography. For membrane proteins, this is still a challenge, but progress is being made. The mechanism of quite a few enzymes is now considered to be understood reasonably well (although fundamental issues remain (Scrutton et al., 1999; Sutcliffe & Scrutton, 2000)) and so are regulatory mechanisms in the sense of which molecule might bind to which other molecule and regulate the activity of the latter. Pathways and networks of metabolism, gene expression and signal transduction have been mapped.

1.4. Cell Biology: The living cell

Near the ~~turn~~ of the twentieth century genomics revolutionized this landscape. This revolution was preceded by a long and ever accelerating progress in biochemistry, molecular biology and the related disciplines of microbiology and biophysics and led to a combined discipline: cell biology. It defined the organization of life at the cellular level in qualitative terms of its molecules. With apologies for the readers who know their cell biology, but with due respect to the philosophers who may not quite do so but are interested in systems biology, we shall now describe the essence of this definition.

Early on, biochemistry had shown that all (most) chemical conversions carried out by living organisms occurred in a number of simpler steps such as dehydration, transfer of phosphate from ATP, dehydrogenation and isomerization. Each of these is catalysed by a protein, called an enzyme, which consists of one or a few chains of amino acids and sometimes an additional organic or inorganic chemical molecule or ion, folded into a complex structure. The amino acids are virtually limited to a set of 20 types. The protein is different for every type of molecule that needs to be converted. This led to the concept of metabolism consisting of large networks of chemical reactions through which mass flows, with a correspondence of every step to a protein (Beadle & Tatum, 1941). The metabolic networks are extremely powerful chemically, being able to convert many types of molecule into many other types, and many thousands of metabolites are known (Kell, 2004). The former correspond to almost anything that occurs in the environment of living organisms and is useful to them as food. The latter are suitable building blocks for the organism. The pluripotency of metabolism appears limited only by impossibilities stemming from a number of fundamental laws, such as the impossibility to create chemical elements from other chemical elements and the impossibility to generate Gibbs-free energy (Westerhoff & van Dam, 1987). The consequence is that there are metabolic networks ensuring that sufficient of each of these commodities is harvested from the food and supplied to biosynthesis. Metabolism is a network that makes biomass from food, although it does not seem to have evolved to be efficient in the thermodynamic sense (Kell et al., 2005; Westerhoff et al., 1983).

The question of how the proteins are synthesized led to the discovery of a network that is orthogonal to this metabolic network at each step of the latter (see Fig. 1). Each protein is synthesized from amino acids by a complex machinery, called the ribosome, which consists of protein and a second main type of macromolecule, i.e. ribosomal ribonucleic acid (rRNA). The diversity of the proteins stems from the fact that the sequence at which amino acids are attached to its nascent chain is specified by a specific messenger RNA (mRNA) molecule. RNA molecules are chains of four types of nucleotide, which are referred to by a mnemonic of the name of the corresponding 'bases', i.e. as A, U, G and C. Each of the 64 triplets of such bases corresponds to an amino acid, with just a few exceptions that deal with the regulation of protein synthesis itself. Each mRNA molecule is a copy of part of single stranded DNA, i.e. a very long chain of nucleotides referred to as dA, dT, dG and dC (the 'd's are often omitted). It occurs in combination with a complementary single stranded DNA molecule which has a T, A, C or G, respectively, where the other strand has an A, T, G and C, respectively. This double strandedness makes the DNA a robust way of storing the information. Damage that can be recognized as such can be repaired by referring to the sequence of the complementary strand. The part of the chain that is copied into an mRNA and is ultimately translated into protein is often called a gene (although this word actually refers to a concept that predates the discovery of DNA). The copying, which is called transcription, is

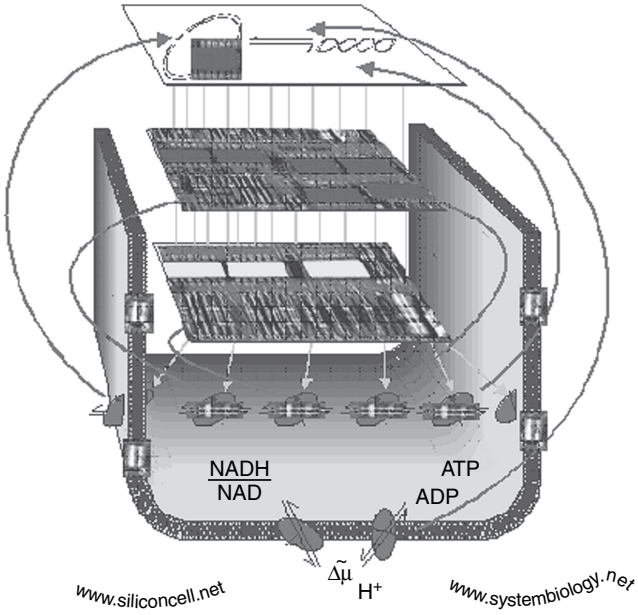

Figure 1 The hierarchical networking of the living cell.

carried out by a large enzyme complex called RNA polymerase. Preceding cell division, the DNA is copied, and the original and the copy end up in different daughter cells.

This set of networks that drive the synthesis of proteins on the basis of information of nucleic acids and information concerning the status of the cell and its environment is one that is often summarized as 'DNA makes RNA makes protein' (see Fig. 1). It is the domain of molecular biology.

Two aspects are of additional importance here: (i) DNA is not converted into RNA, nor is RNA converted into protein. This is a difference with a metabolic pathway where material parts ('mass') of the first molecule ends up in the last molecule. The gene-expression pathways only transfer information. (ii) Where the scheme suggests a hierarchy, DNA directing RNA, which directs enzymes, which then catalyse and hence also direct metabolism, this 'hierarchy' is not dictatorial but 'democratic' (Westerhoff et al., 1990): The rate at which transcription occurs depends on the binding of other proteins (called transcription factors) to parts of the DNA close to or relating to the gene. That binding in turns depends on the concentrations of metabolites that may bind to these, depending on whether the transcription factors are in the proximity of the DNA or depending on whether they have been modified chemically.

The chemical modification of transcription factors responds to the status of intracellular metabolism and to the presence of extracellular signals, such as light, and the presence of food. This response is achieved by yet another set of networks. These networks specialize in this signal transduction and again consist of pathways in which each step is catalysed by proteins. In most of these pathways however, there is no transfer of mass from the beginning to the end. Only information about the conditions measured at the beginning of the pathway is reflected by the state elsewhere in the pathway.

Metabolism, gene-expression and signal-transduction constitute networks in the dimensions of time, information and chemistry. The living cell also depends on other networks that address the dimensions of chemistry, structure and space. The cell itself is a membrane-bounded compartment. In eukaryotes such as mammals, the cell also contains many membrane bounded subcompartments, which house networks that can be incompatible with networks in other subcompartments. Without catalysis, transport across most of the membranes is impossible, and the transport of some macromolecules through compartments is also catalysed. The DNA is folded into a complex structure with proteins called chromatin. These networks of structure and transport through and around structures have been well characterized. In recent years, more and more of these structures have been shown to be displaced from equilibrium, being maintained continuously by regulated active networks. Examples include the DNA structure, certainly in bacteria (Snoep et al., 2002), the asymmetric lipid distribution in membranes and the microtubular and actin networks in the cell sap.

Molecular biology became a further success story when it joined forces with biochemistry and microbiology and became modern biotechnology. First, it was discovered that many organisms make enzymes that cut DNA with specific nucleotide sequences. By not having those base sequences themselves, those organisms could protect themselves against invading viruses. These 'restriction enzymes' were used by scientists to put genes of interest into organisms that did contain those sequences. By growing these organisms and then again applying the restriction enzyme to isolated DNA, pieces of DNA corresponding to genes could be 'cloned', i.e. purified and their amounts greatly amplified. The resulting material could then be introduced into other living cells which would then express that DNA into protein. If those cells altered their functioning this helped establishing the function of the gene. The amplified amount of DNA also enabled that DNA to be sequenced, first by tedious methodology, but in a demand-driven mode this led to the development of new and rapid sequencing methodology. (The methodology to amplify DNA also became much more effective when the polymerase chain reaction (PCR) was developed, allowing for the amplification to occur in vitro.) The result was that the nucleotide sequence of each gene of interest could be determined. Because of the 64-to-20 mapping of DNA sequence to protein sequence, this implied that the amino acid sequence of the corresponding protein was also determined. Through the above cloning procedure, larger amounts of proteins could be obtained enabling structure determination through X-ray crystallography and NMR. At present the structure of almost any soluble protein can be determined, albeit at relatively low throughput.

It also became possible to determine whether any given gene was expressed in an organism. Here the base-pairing phenomenon that underpins DNA and mRNA function served molecular biology. Tagged DNA or RNA molecules that were complementary in terms of nucleotide sequence were synthesized and made to react ('hybridize') with mRNA isolated from living organisms. If a certain mRNA was expressed then the hybridization would betray this. Because so many genes are expressed in any organism and because of background reactivity, the mRNAs first had to be separated from each other, which was accomplished by gel electrophoresis. A corresponding methodology was developed for the measurement of expression at the level of protein, by using specific antibodies for the proteins. The separation power of these methods is however limited, and therefore they were not suitable for genome wide measuring of gene expression.

Another powerful tool came from genetics applied to rapidly growing microorganisms. Mutations were made in the DNA of these organisms and the consequences for their functioning was determined. Through the above methodologies, mutations could be related to proteins. Deleting genes and observing the consequences, pathways could be constructed that should be responsible for certain aspects of cellular behaviour. When this was done for different organisms, an astonishing phenomenon turned up. This was the extensive homology

of organisms in terms of their intracellular organization, as well as in terms of the amino acid sequences of their corresponding proteins. In principle, the major food substance glucose could be oxidized in many ways to carbon dioxide with the harvest of much of the corresponding free energy. Virtually all organisms, however, possess the glycolytic pathway and the tricarboxylic acid cycle, and many contain the membrane-associated electron-transfer chain, which comprise one way of accomplishing this overall process. *A fortiori*, the enzyme that catalyses the phosphorylation of glucose by ATP, is sufficiently homologous also in terms of its amino acid sequence, for its sequence to be identified in many newly sequenced genomes, through the sophisticated techniques of bioinformatics. Even more strongly so, functional domains of proteins (such as ATP binding sites) have been sufficiently conserved through evolution to be recognized between genomes. On another planet with perhaps much higher rates of net mutagenesis, and much lower selection pressure, this may be different, but for our planet this phenomenon of extensive homology has been an enormous asset to molecular biology. To many newly sequenced genes, a function is assigned simply on the basis of sequence of homology, and in many cases this assignment turns out to be correct, qualitatively. An important consequence is also that the phrase 'understanding life' does have a meaning. It could have been such that molecules, mechanisms and pathways differed immensely between organisms and that each organism had solved the problem of how to stay alive in its own, entirely different way. It is quite clear now that this is not the case; life as we know it in a broad sense is probably maintained in just one single way, with 'minor' variations on the theme. This is not to say that this variation, which is minor in terms of principle and quality, is not vast in terms of quantity. Biological diversity especially in the microbial realm is enormous. Accordingly, life is able to maintain itself under a very wide variety of conditions on this planet, but again, essentially through extensive variation on a single theme. Of course, this greatly motivates the scientific question of what constitutes this essentially uniform molecular basis of life.

The maps and structures of living cells, i.e. the field that may be called cell biology, were considered known in the 1980s in their essence. What was lacking was the completeness. Although for each type of network, a number of examples had been well documented, many actual networks had not yet been identified. More disturbingly, however, every now and then a cellular component was discovered that was strongly involved in the already 'known' pathways, most often in their regulation, but often even in their mechanism. Examples included fructose 2,6 bisphosphate in glycolysis, the chaperonins in the protein synthesis pathway and ubiquitinylation in signal transduction. In addition, although some cellular behaviour could be explained qualitatively on the basis of the known networks, much other behaviour was in conflict with what was known, or simply not explained by it. The conflicts could not be used constructively as falsifications

(Popper, 1992), because it was well recognized that there were many unknown components and regulatory mechanisms in the cell that could affect the pathway that was under investigation. For similar reasons verifications were limited in value. What often resulted was an escape of biochemistry and molecular biology to well defined in vitro systems, where at least the mechanisms of the proposed pathway or molecules could be established, even though the relevance for their operation in vivo became unclear.

2. LIMITATIONS TO THE SCIENTIFIC STATUS OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

Notwithstanding their success concerning the understanding of single types of macromolecules, classical biochemistry and molecular biology face limitations when compared to the science aimed at by the philosophers of classical physics. These limitations are

(1) Inaccuracy: no quantitative, i.e. accurate testing of hypotheses
(2) Inability to deal with emergent properties: because of lack of quantization it is impossible to test a number of qualitative hypotheses that are highly important for the emergent properties in living systems
(3) Irreducibility: biochemistry and molecular biology theories cannot be reduced to physical chemical theories
(4) Impotency, i.e. inability to address Life itself and lack of connection to organismal Biology
(5) Undefinedness: not all factors that play important roles are known and consequently hypotheses cannot be tested
(6) Inaccessibility to experimentation: the systems under study cannot be experimented on through a sufficient number of degrees of freedom
(7) Lack of analyzability

We now discuss these limitations, one at a time.

2.1. Inaccuracy

The first limitation is that the cartoon-type hypotheses were not quantitative and thereby unfit for the strictest possible quantitative testing, a procedure desired by the philosophy of physics (Carnap, 1966). Being quantitative enables tests to be more stringent (Laughlin, 2005). If the temperature of a closed vessel with an ideal gas rises by 10% then the qualitative test of the law of Boyle asks if the pressure goes up, whilst the quantitative test asks whether the pressure goes up by precisely 10%. Clearly, the qualitative test has a 50% chance of being passed

by coincidence, whereas the quantitative test has a much smaller such chance, depending on the experimental accuracy.

2.2. Inability to deal with emergence

A second limitation also derives from the lack of being quantitative but, paradoxically, pertains to failure to test the prediction of qualitative phenomena. The behaviour of systems of independent components is nothing but the simple addition of the behaviour of those components. In sufficiently nonlinear systems and even in linear systems with certain networking (for simplicity we shall here call the latter also 'nonlinear'), qualitatively new behaviour may emerge, which is often important for biological function. In fact for survival of living organisms, a number of properties is essential that are absent from the individual molecules in those organisms. They must emerge from certain nonlinear interactions. We shall refer to those nonlinear interactions as 'essential' nonlinearities. Examples include oscillations in networks of components that would themselves never oscillate (Goldbeter et al., 2001), and free-energy transduction between components that would themselves only dissipate free energy (Westerhoff & van Dam, 1987). For biological macromolecules, the nonlinearity varies between conditions, as it depends on their environment. We briefly illustrate this by considering what may be the rate equation of an enzyme in an intracellular network:

$$v = \frac{[S] \cdot V}{K_m + [S]} \tag{1}$$

where v, [S], K_m and V refer to the actual reaction rate, the concentration of the substrate of the reaction, the Michaelis–Menten 'constant' and the 'maximum' reaction rate, respectively. The way in which the enzyme affects the behaviour (both in the qualitative and in the quantitative sense) of the network is fairly well described by the elasticity coefficients for the metabolites with which it interacts, in this simplest case, the substrate S. This elasticity coefficient corresponds to the log–log derivative of the rate with respect to the concentration of the substrate, i.e.

$$\varepsilon_S = \frac{\partial \ln v}{\partial \ln [S]} = \frac{K_m}{K_m + [S]} \tag{2}$$

The equation shows that the role of the enzyme in the system is not only determined by that enzyme itself (through K_m) but also by its environment (i.e. by $[S]$) and by how it interacts with that environment (in terms of S/K_m).

Whether the new behaviour that emerges depends on the type of nonlinearity that reigns in the network, e.g. on the value of the above elasticity coefficient. Consequently any theory explaining the occurrence of oscillations will only

predict oscillations for certain states of the system (i.e. certain magnitudes of [*S*]) and not for others (as is observed, e.g., Ihekwaba et al., 2004; Nelson et al., 2004), and their nature can depend even qualitatively on multiple enzymes in the system (e.g. Ihekwaba et al., 2005). Testing whether the theory indeed explains oscillations that occur in a living cell will first have to determine what the state of the system is, in a quantitative sense (i.e. how high [*S*] is, and not just whether there is some S or not), then to ask whether for that state the theory predicts oscillations, and then to test whether under those conditions oscillations are indeed observed experimentally. The implication is not only that theory and experiments need to be quantitative but also that they need to pertain to the conditions of the living state, i.e. they need to be performed under conditions as close as possible to those that are considered to pertain in vivo, preferably in the living organism itself.

An actual example is the following. If one observes synchronous glycolytic oscillations in intact yeast cells (Davey et al., 1996; Richard et al., 1993), and one proposes that the stimulation of the enzyme phosphofructokinase by AMP is 'responsible', one can test this hypothesis by mutating the enzyme and removing that stimulation. However, any alteration that alters the system such that its state is no longer in the oscillatory domain, will do away with the oscillations. In fact the proposed mutation of phosphofructokinase could well do away with the oscillations by simply shifting the system to a different operating point even if this product stimulation were responsible for the oscillations. A proper test of the hypothesis thus removes the AMP effect whilst simultaneously modulating the system so as to keep it at its operational state. Better, one removes the AMP effect gradually and asks if the frequency or amplitude of the oscillations changes (Reijenga et al., 2005b). In nonlinear systems, even qualitative statements therefore need quantitative tests (Ihekwaba et al., 2005).

How important is this issue? Well, the rate and equilibrium equations for most biological processes are nonlinear or at least nonproportional (Hill, 1977; Westerhoff & van Dam, 1987). Moreover, many of the biological processes that are important for function exhibit properties that one would not see in individual molecules and that therefore require nonlinear interactions between those molecules. These processes include differentiation, development, the cell cycle, robust signal transduction and most transport processes. Their theories can only be tested if they are quantitative, and strictly only by quantitative experimentation that is performed inside the living cell.

2.3. Frustrated aspiration of biochemistry and molecular biology to . . . biology

Another type of limitation to biochemistry and molecular biology is that they do not by themselves produce the overlying science, i.e. biology. In principle,

biochemistry and molecular biology study all the molecules that occur in organisms, but they refrain from addressing the life that is embodied by all those molecules. Although this claim of insufficiency of biochemistry and molecular biology has often been made by physiologists and other organismal biologists, it is not immediately appreciated by all, and certain its remedy is not. There is indeed a paradox: if biochemistry and molecular biology were to continue to study and establish the structure and the mechanisms of action of every macromolecule of a living organism, then they should ultimately understand that whole living cell. For what else is there in a living cell than its molecules? This contention is the most common version of the reductionist agenda: dissect any system into its elements, study all those elements individually, and then just understand the system. Technically, the 'just understand the system', implies that systems behaviour can be understood as a superposition of how all its components behave individually. The organismal biologists often observe that when a living system is taken apart, it loses much of the essential behaviour of living systems. This makes some of them turn to the holist agenda, which studies only intact systems. This then makes them subject to much of the limitations noted above for organismal biology, and more importantly, it implies that those limitations will stay forever, independently of the progress of science.

The reductionist and holist paradigms seem to be irreconcilable, but below we shall propose that through systems biology and the silicon-cell approach they may not be. Here we shall first indicate why the 'just understand the system' methodology does not work, i.e. why by themselves biochemistry and molecular biology cannot produce biology. The reason is again the essential nonlinearities of biological systems. Much of biology depends on dynamic phenomena that emerge in nonlinear interactions. These cannot be understood by the simple addition of the behaviour of the components in isolation. This is one reason why biology lies outside the realm of biochemistry and molecular biology *sensu stricto*. In other words, what makes a system different from its parts list is the non-linear interactions between those parts, and these are changed or lost upon disassembly.

2.4. Irreducibility

A third limitation is again related to the cartoon aspect of biochemistry and molecular biology: in these new disciplines molecules are not drawn in terms of their structure or chemical equation, but by coloured balls with short, non-chemical names, such as hexokinase, HXK, Ras or wnt. These names serve reasonably well as mnemonics. Attempts to give enzymes systematic names produced names that referred to their activity rather than to their chemical formula or structure. The reason was that for many enzymes the chemical structure could not be established, whereas at least some of the catalytic activities could be. The

names and concepts of biochemistry were not reduced to the underlying physical chemistry (in the sense of reduction of theories to underlying more general theories, cf. above). Similarly, 'the' structure of nucleic acids and proteins was determined by X-ray crystallography, but the question of whether that structure was stable with respect to the physical forces between amino acids and between bases, was not addressed. This was in part because it could not be addressed effectively. Virtually none of these structures can presently be calculated ab initio (see (Popelier & Joubert, 2002) for an example), precisely because the interactions are nonlinear, and with many interactions depending on other interactions. Likewise, electric field effects on transmembrane movements of ions cannot be vested in physics and chemistry because too much of the details of the transport matters and is in fact unknown. Although there has been some progress in the calculation of enzyme catalysis in terms of physical–chemical interactions, most such reaction mechanisms cannot be verified in terms of precise physics and chemistry. The same is true for the pathways of processes that make living cells operate. The fluxes through them cannot be calculated ab initio either, but only from direct physical–chemical interactions and atomic structures. In biochemical textbooks, pathways are therefore drawn as roadmaps running through many towns and connecting major cities or hubs (Barabási & Oltvai, 2004). Indeed, reduction of molecular biology and biochemistry to the underlying physics and chemistry is rare, and not even an aim of these disciplines anymore; both disciplines are entirely successful on the basis of their own concepts and laws, immaterial whether these are reducible to physics and chemistry or not. However, this general problem of intractability in terms of the underlying physics and chemistry caused reluctance among many physicists and chemists to consider biochemistry and molecular biology as serious sciences. The biology of entire living systems was observed to be too complex and ill defined for the hypotheses to be strict, testable and falsifiable. To some, this made molecular biology and biochemistry appear to remain stamp collecting.

Indeed, the above limitations suggest that neither biochemistry nor molecular biology connect to physics. They fail to meet the criteria of classical physics that were once proposed to be the criteria of proper science (Carnap, 1966). Looking at chemistry beyond quantum chemistry, this may not be a novelty among the experimental sciences; chemistry may not connect to physics either (Primas, 1981).

2.5. Lack of testability because of undefinedness

Another important limitation of biochemistry and molecular biology relates more literally to holism. Returning to Eqn (1), we realize that the Michaelis 'constant' is independent of the concentration of S but not necessarily constant otherwise. Agents binding to the enzyme catalysing the reaction may influence

this Michaelis constant, and certainly the concentration of the product of the reaction changes the (effective) K_m, i.e.

$$K_{m,\,\text{apparent}} = K_{m'} \cdot \left(1 + [P] \Big/ K_p\right) \tag{3}$$

All components of the same living cell may therewith affect the role the enzyme plays in the cell's behaviour, ~~also~~ the components that are not yet known .... This pinpoints one of the arguments of holism, in that to understand the role of one of the molecules in a system with the type of nonlinearities found in cell biology, one must look at the whole. We do not think that one should necessarily be able to look at the whole all the time, but certainly one should be able to look at all the possible molecular factors that play a role. Until recently, not all molecules of the living cell were known or even knowable, making it impossible for biochemistry and molecular biology even to determine with certainty the role a molecule of choice might play in determining the nonlinear behaviour of the living system, simply because unknown factors could well be clouding any issue. Post-genomics is beginning to change this.

2.6. Lack of experimental accessibility

As emphasized by Carnap (1966) for physics, it is important that hypotheses are tested under all relevant conditions and in terms of all relevant degrees of freedom. In living systems, many factors may exert an influence and it should therefore be mandatory that proposed mechanisms are tested by modulation of all those factors individually. For as long as not all those factors were known, it was difficult for biology to carry out these tests; the living system was not accessible enough for such testing.

2.7. Lack of analysability

Because many factors are likely to be involved in the sustenance of the living state, hypotheses concerning mechanisms are likely to be multifactorial. Accordingly, many of these factors should be monitored simultaneously in tests. Although quite a few factors can be measured individually by biochemistry and molecular biology, until recently it was impossible to monitor many components simultaneously.

Summarizing, we see a landscape where biochemistry and molecular biology could extend neither to physics nor to organismal biology because of at least these seven types of limitation. We shall now discuss recent changes in molecular biology that would seem to do away with some of these limitations.

3. RISING ABOVE THE LIMITATIONS

3.1. Genomics

A major cause of the above limitations was that there existed no complete understanding of inventory of all the components of a living cell, even though such an inventory had been identified in principle, i.e. the DNA: the DNA contains the information for all the proteins in the cell and the proteins catalyse all the reactions. It was thought that in principle, the sequence of the DNA should determine everything that happens in the living cell, under any given set of environmental conditions. It became quite important therefore to sequence all the DNA of a living organism, and in the 1990s of the previous century, large consortia of researchers embarked on accomplishing this aim in activities referred to as 'genomics'. It may seem that genomics was not much different from the molecular biology that preceded it. Indeed, many of the most active scientists in genomics continued to be molecular biologists as well. Yet, for our discussion here, the transition between molecular biology and genomics has been quintessential. Genomics went after the determination of the complete DNA sequence of an organism, rather than of DNA sequence of many of its components, i.e. genomics went for the system rather than for its components. By 1995, the first complete sequences of the genomes of free-living organisms (cf mitochondria in 1981 (Anderson et al., 1981)) became available (Fleischmann et al., 1995), and importantly also the sequences of the two best-known model organisms soon followed, i.e. the eukaryote yeast (Goffeau et al., 1996) and the bacterium Escherichia Coli (*E. coli*) (Blattner et al., 1997). By 2001, the DNA sequence of humans was nominally established and sequences of many organisms have become known as we write this. In essence, the DNA sequence of any organism can now be determined. Because of the homology discussed above and thanks to bioinformatics, the function of many genes can be proposed with appreciable success rates when the homology to genes of known function is close. Although for half of all sequenced genes (this fraction differs between organisms), the function is uncertain or unclear, this fraction is considered to be on the decrease. (We would stress, of course, that many genes with some 'known' functions will turn out to have other functions that are as yet unknown.)

Knowing most of the genes of an organism provided a strong motivation for what has been called 'functional genomics', i.e. for determining whether those genes function in terms of being expressed and what their role is. Because of the strong tendency of nucleic acids of complementary sequence to react with each other, this was possible in principle by making populations of small RNA molecules each of which was complementary to part of one of all the genes in the genome. A breakthrough came when those probe molecules could be spotted as an array onto a slide and could be provided with a fluorescent tag that lights up when an mRNA molecule hybridized. This nucleotide array technology is

now used to determine the expression of all genes at the level of mRNA, at accuracies beyond 30%.

No similar hybridization chemistry exists at the level of a chain of amino acids (yet). Using immunological techniques however, antibody-like molecules are now spotted onto arrays, and the abundance of proteins in extracts from cells is determined (Walter et al., 2000). Alternative modes of genome-wide detection of protein abundances include a methodology in which all proteins are separated in a highly reproducible way through two-dimensional (2D) gel electrophoresis, such that each location in 2D corresponds to a specific protein. The mapping of spot location to the identity of the gene is a slow process, but for smaller genomes this methodology is getting close to the possibility of genome wide detection of gene expression at the level of protein. This methodology is inherently limited in three important ways. First, the resolution of 2D gel electrophoresis is insufficient to separate all proteins of genomes larger than a few thousand genes; though useful for bacteria, the methodology is still of more limited value for human biology. Second, the method is not quantitative yet, and indeed many proteins, especially membrane proteins, are missed entirely. And third, it is difficult to identify the individual proteins. The latter problem is now being alleviated by the implementation of mass spectrometry. By extracting protein from a specific location on the 2D gel, subjecting that to limited proteolytic digestion, determining the precise mass and/or sequence of the resulting peptides and combining the resulting information with the known sequence of the genome, the protein spots can now often be attributed to specific proteins.

Mass spectrometry also offers methods that may analyse genome-wide expression at the protein level. The gel-electrophoresis step can be replaced by capillary chromatography, a separation by mass spectrometry on the basis of the total mass of the protein (or fragments thereof), fission of the protein/peptide in the gas phase and then a second mass spectrometry step to determine what the resulting fractions are. Again the availability of the genome sequence enables one to identify the protein. For mass spectrometry, molecules have to be brought into the gas phase as electrically charged molecules. However, existence in the gas phase is far from the thermodynamically most favourable mode of existence for most of the molecules that constitute the living cell. The effectiveness at which the entry into the gas phase is achieved is low therefore more importantly, it depends much on the presence and properties of the other molecules in the mixture. Other molecules with electric charge can affect the tendency of a given molecule to enter the gas phase. Consequently, the mass spectrometry method is inherently irreproducible in the quantitative sense; it is hard to determine expression levels accurately with this method (although this is improving both by changing conditions in the mass spectrometer (Vaidyanathan et al., 2003) and by isotope-based quantification. This is because isotopes behave essentially identically with respect to the above problems, yet can be discriminated readily

by the mass spectrometer. Spiking samples with known amounts of an isotope of the substance of which the quantity needs to be determined, therefore enables quantitative determination of amounts of proteins (more often in relative terms but occasionally absolutely (e.g. Beynon et al., 2005).

The genome-wide determination of gene expression at the levels of mRNA and protein are called transcriptomics and proteomics, respectively. Genome-wide analysis of the expression at the level of metabolism, which is often closest to function, is called metabolomics. Genome-wide metabolomics has not yet been developed to the same extent as transcriptomics (Dunn, Bailey & Johnson, 2005; Dunn & Ellis, 2005; Goodacre et al., 2004). Mass spectrometry methods akin to the ones described above for proteins are being developed for metabolomics. Again it is a problem to get the metabolites into the gas phase and to determine their level quantitatively; isotope methodology can again solve this problem (though one needs an isotope for each determinand, and the larger problem resides in the fact that we do not know what most of these molecules are . . .).

Cell function is determined not only by the expression levels of proteins but also by where they are expressed. Here three developments are highly important. One is that of high-resolution microscopy. The second is the development of many fluorescent probes for important molecules and ions in living cells. And third is the possibility of fluorescence- or luminescence-based reporter proteins, which are either fused to proteins of interest or are put under the control of the gene-expression control elements that normally drive the expression of proteins of interest. Thanks to these methodologies, the timing of expression and the dynamic localization of many molecules in the living cell can now be determined.

Another less profound, yet highly important advance in technology is that of robotization and automation for high throughput experimentation. By using plates with many reaction vessels and robots doing the pipetting, many experiments can be performed in parallel and at much enhanced reproducibility.

At present one can determine for all genes in a genome simultaneously whether they are expressed at the level of mRNA. Soon this will also be possible at the level of protein and in terms of their relationship to further levels of functionality, e.g. at the level of metabolites. Through functional genomics, therefore, everything will potentially soon be knowable and known about living cells. For unicellular organisms this should imply that everything will be known about a living organism, albeit that collections of such cells remain highly heterogeneous (Davey & Kell 1996). Every component can be manipulated by expressing the corresponding gene in the organism under the control of a regulatory element that can be steered by the experimenter. Everything will come to be known therefore and systems of Life will come under complete experimental control. The limitations of the 'undefinedness' and inaccessibility to falsification–verification experiments of biology, will soon be gone. Finally biology can stop collecting stamps and become 'proper Physics', or so it would seem.

3.2. Soon everything will be known . . . : Will biology become physics, at last?

Indeed, the vast increase in power of molecular biology, and the ability to experiment and analyse genome wide, should get biology much closer to the ideal of constructing completely verifiable and falsifiable theories. Of the above list of seven limitations, it would seem that the ones regarding undefinedness, inaccessibility and lack of analysability have disappeared with the advent of functional genomics. These three criteria come close to the criteria that proper physics should adhere to, e.g. according to Carnap (1966). Provided that the analyses of functional genomics are made quantitative, it would seem that the first criterion (accuracy) will also be met. It would seem therefore that with functional genomics Biology would all but graduate to become proper physics.

From the point of view that science should be one and indivisible, the reduction of biology to just another physical chemical science with 'just' the same methodologies and quality criteria, would seem to be a great good. Whether this should actually happen is the fundamental issue that is the subject matter of this book. We shall now indicate why we think that this reduction is not to be expected.

3.3. Observing or understanding?

Functional genomics will enable us to observe virtually everything that happens in living organisms. The aim of the sciences, however, is also to understand the observations. Such understanding can consist of the possibility of deducing what is observed from pre-existing theories. It can also amount to the understanding on the basis of theories that are being generated as many more observations are made, i.e., through induction, principled hypothesis formulation and hypothesis testing through verification/falsification procedures.

We shall first address the former basis of understanding. It turns out that functional genomics has not removed the limitation of irreducibility from biochemistry and molecular biology, and that it will not do this in any foreseeable future. When it was proposed to sequence the whole genome of organisms, one of the underlying arguments might have been that this should automatically lead to the understanding of the functioning of living cells and organisms in molecular terms. Folding of a protein was perhaps thought to be determined by it finding the structure with the lowest free energy. Because that free energy is determined by the interactions of all its amino acids and the sequence of these in the chain, it was perhaps thought that one should be able to calculate that structure ab [illegible]. For all but the simplest proteins, the calculation of the structure with the lowest free energy from the amino acid sequence is still impossible. The problem is strongly nonlinear and hence much too complex to be carried out by existing

computers. In fact, the calculations of protein structure that are being done with some success are not truly ab initio but use phenomenological force fields and/or knowledge of existing structures. At present structure predictions of proteins on the basis of their sequence are occasionally fairly successful, but such predictions are virtually only based on comparison with homologous structures. The next step, i.e. the calculation of catalytic action from the protein structure is equally difficult. Here too, success is based on comparison of homologous series. The ab initio calculation of kinetic properties of entire pathways might all be possible in principle, but it is impracticable at present and in fact for any foreseeable future, due to the sheer complexity and nonlinearities of the interactions that are involved (see also Westerhoff & Kell, 1987).

In the living cell there are also catalysts of correct protein folding, i.e. chaperonins or by the action of the ribosome. Because both these assisting proteins couple this process to a reaction consuming free energy, it is quite possible that they put their target protein in a structural state with a free-energy that is higher than minimal. Indeed, the structure of proteins may not even correspond to the free energy minimum but be determined by the mechanism of folding. After all, the spontaneous conversion between native and denatured states of proteins is rarely effective.

A lingering feature of biology could well be important here. This is its inherent hysteresis. The concept of biology as straightforward though complicated physical chemistry, should be most consistent with the following picture of the genesis of a new living cell: in an existing living cell all the components of a daughter cell might be synthesized independently de novo, inclusive of the lipids necessary for its membrane and its DNA. Then a closed spherical lipid bilayer would be formed around all the newly synthesized components, and the newly formed cell that sat inside the mother cell would be extruded by that mother cell. After their synthesis, all components for the new cell would assume their minimum free energy structure independent of the activities of the mother cell. The state of the daughter cell would then be determined entirely by free-energy minima, hence by the physical chemistry of its molecules. This mechanism of generating new cells might be entirely possible and would in fact be consistent with what Van Leeuwenhoek expected to see in terms of *homunculi* through his microscope. But it is not what actually happens. Instead, the membrane of the daughter cell is formed by splitting off a part of the membrane of the mother cell; the DNA of the daughter cell is the result of a semiconservative replication of the mother cell, i.e. the mother and the daughter cell receive both one strand of the DNA of the mother cell, the other strand having been synthesized de novo. According to our current knowledge, the proteins that end up in the daughter cell are not all proteins that have been synthesized de novo. Newly synthesized proteins and pre-existing proteins and even newly synthesized organelles and pre-existing ones end up in both the daughter cell and the mother cell. In many

organisms the mother cell after division and the daughter cell are effectively the same; division yields two daughters and the mother ceases to exist. In other organisms such as *Saccharomyces cerevisiae*, division is asymmetric, and the mother differs from the daughter, yet appreciable mixing has occurred. Importantly also, the DNA, mRNA and proteins of the young daughter cell have been synthesized by the DNA polymerase, RNA polymerase and ribosomes of the mother cell. Consequently, rather than that each cell is an entirely new physical-chemical phenomena, all cells are in fact continuous with each other. If there were a process of excessively slow relaxation in a cell, the same process would be in the same state in all daughter cells. That this not so is in part is reflected by observations of epigenetic phenomena.

The extent to which this possible hysteresis is actually important is unclear at the moment. For molecules of low molecular weight and complexity, it is unimportant because relaxation to an equilibrium structure is fast enough. For macromolecules and for the regulatory state of networks it might be important. This issue simply has not been looked at sufficiently yet. In some cases of regulation, such as for instance with the *lac* operon in *E. coli*, the regulatory state is effectively inheritable through this type of mechanisms, which has the effect of zonation of its colonies. In its ultimate form the point of hysteresis appears obvious. All amino acids in proteins have the *L*-stereoisomeric constellation. The mirror world with all *R* amino acids should be energetically equally probable. Yet new cells with all their proteins in the *R* form do not arise, because the enzymes that make their amino acids make the *L* form.

The conclusion is that the feature that it is too difficult to calculate structures of proteins on the basis of physical–chemical principles may not even be too relevant. It is quite possible that most of the structures that exist in living cells are determined by more than the straightforward physical chemistry of those molecules themselves. They may also depend on pre-existing structures of other molecules with which they interacted during synthesis. The fact then that biochemistry and molecular biology do not start from underlying physical–chemical principles but with their own elementary objects such as enzymes and genes, may be an asset rather than a disadvantage. The corollary is that also the irreducibility of biochemistry and molecular biology to physics is much more fundamental than technical. Any molecule-based biology may therewith be a science that is fundamentally different from physics.

Evolution has not selected structures with maximum entropy (Schrödinger, 1944), minimum free energy (Nicolis & Prigogine, 1977) or maximum thermodynamic efficiency (Westerhoff & van Dam, 1987), and in fact much of the functioning of biological replication may have been structured so as to prevent relaxation to such a state. Also here simple physical–chemical considerations do not suffice to understand biology. As also proposed by Schrödinger (1944), biology warrants its own explanatory principles.

Of course physics too has undergone a tremendous evolution since the days of Schrödinger and Carnap (Schrodinger, 1944; Carnap, 1966). It has been recognized that far away from equilibrium, physical–chemical systems may relax towards metastable states rather than to equilibrium, and anyway such states are typically well isolated from each other in the form of local minima as in any other search landscape (Bäck et al., 1997; Frauenfelder & McMahon, 2001). The states can be more complex than the equilibrium state, i.e. appear to be more organized than the latter. Such physical self-organizing systems have been proposed to be at the basis of the tremendous organization that is observed in biology. Accordingly, parts of modern physics address the generation and maintenance of complex dynamic structures, and how new properties may emerge from nonlinear dynamic interactions. However the mechanisms that have been proposed such as the Brussellator (Nicolis & Prigogine, 1977) are themselves nonverifiable/nonfalsifiable. This is because they were formulated in much too general terms, causing loss of the specificity of the biological system at hand. Testing of nonlinear phenomena requires precision, hence a precise matching of mathematical model and experimental system. Wolf et al. (Wolf et al., 2000) have recently worked towards such a testing of a proposed self-organization mechanisms for synchronization of the glycolytic oscillations in a population of yeast cells, but this may only serve as an incomplete example. This brings us to the second type of understanding, i.e. on the basis not of the principles of underlying sciences but of principles that are discovered in the science at hand, i.e. on the basis of newly discovered principles of biological systems. Here there is the issue whether anything is to be expected from the search ~~of~~ such theories.

Metabolic and hierarchical control analysis are theories that may serve as examples of theories that are custom-made for biological systems (Westerhoff & Hofmeyr, 2005). By making an idealized description of intracellular networks, i.e. metabolic networks for the former theory and gene-expression or signal transduction networks for the latter, a mathematical set of definitions can be made and laws can be deduced from the time-transformation invariance and from stability against fluctuations (cf. Hornberg et al., 2005; Peletier et al., 2003; Westerhoff & van Dam, 1987). These theories are in a sense comparable to theories in physics in that they derive from observations that falsified alternative hypotheses, and led to conjectured new laws, which could then be deduced from postulated fundamental properties (axioms) of the system. Other 'laws' that derive more as a result of induction from experimental observations may also be found for biological systems, such that proteins are encoded by mRNAs which are in turn encoded by pieces of DNA, and the law that for every natural substance on this planet that can be broken down to yield free energy, there is an organism that does precisely that.

On the basis of this experience, we expect that many more theories will be established for living systems. These will differ from those we already know

from physical chemistry and then not only in terms of their precise meaning, but perhaps also in terms of their structure. Perhaps such biological theories will be less general, more condition dependent, and much more complex. This remains to be seen. Automated hypothesis generation from experimental data may show new ways in this respect (King et al., 2004).

3.4. Systems biology

Our contention that the molecular biology of living systems is neither physics nor biology, but rather a science in its own right, suggests that it is entitled to a name. Such names already exist, i.e. systems biology and integrative biology. We shall here use the former. We propose that systems biology attempts to establish principles of operation of biological systems such as the living cell. It should thereby find its own concepts rather than reduce them to physical chemistry. It should strive to be quantitative enough to be able to understand the emergence of functional properties from nonlinear interactions between components of biological systems. It should also appreciate that such interactions depend on the precise state that the biological system is in. This has the consequence that laws should address specific conditions rather than be completely general. For instance a law could be that the glycolytic pathway can engage in oscillations provided that the elasticities of the following stated reactions fall within the following range . . . . The law should not be of the generality of physics i.e., that the glycolytic pathway might engage in oscillations under any, unspecified conditions. Systems biology should synthesize the following features

(1) Information on expression levels is contained in the DNA and is expressed through mRNA into proteins which then catalyse reactions.
(2) The expression levels are not simply determined by transcription activities of the DNA in a dictatorially hierarchical fashion, but controlled by a combination of extracellular signals and intracellular concentrations.
(3) The concentrations of intracellular substances are determined by all the intracellular processes and extracellular together.
(4) The intracellular processes are determined by the expression levels of the enzymes, by the kinetic parameters of those enzymes, as well as by extracellular signals and intracellular concentrations.
(5) Much of biological regulation is one of circular or spiraling causality (Rosen, 1991; Westerhoff & Hofmeyr, 2005), i.e. a concentration of a substance may co-determine the concentration of another substance at later times and be co-determined by the concentration of that other substance at earlier times.

(6) Due to nonlinear interactions, qualitatively new properties may emerge; whether this happens depends on the precise magnitudes of the parameter values.
(7) Part of the structure and dynamics of the living cell may be prespecified by evolution, by its mother cell and by the synthetic machinery therein.
(8) Living organisms are the product of dynamic interactions between structures and chemical reactions, where the latter determine the former and the former determine the latter to quite significant extents.
(9) Much of biological mechanism and regulation is not determined by any single factor but by a multitude of factors.
(10) The simplicity of mechanisms that serves as Occam's razor in the decision between competing theories in physics is of comparatively lower real value in biology. Functionality and fitness and empirical facts rule over simplicity. The actual mechanisms in systems biology may be more complex than possible because of coselection for other purposes in evolutionary optimization, because evolution may have led to systems that are optimal locally but not globally, and because simplici[illegible] human eyes may be complex in systems biology terms (and vice versa).

Much of life is associated with organizational and intelligence aspects that 'emerge' from molecular behaviour (Kell & Welch, 1991). Although these emergent properties are not in conflict with physics and chemistry, much of physics and chemistry traditionally shies away from complexity, hysteresis and nonlinearity (although other parts such as those dealing with superconductivity, lasers, ferroelectricity and other highly nonlinear phenomena cannot escape it). As we discussed above, their paradigms favour the kind of simplicity and Occam's razor strategy that may not be relevant for biology. We propose that this makes systems biology (the part of biology that focuses on this kind of complexity) its own science with, indeed, its own methodology and its own philosophical foundations. We shall here then seek to contribute to the development of a philosophical basis for this new science by describing some of the modes in which it operates in practice.

4. TOWARDS A SYSTEMATIC METHODOLOGY OF SYSTEMS BIOLOGY

Other chapters in this book describe philosophical aspects that underlie modern systems biology. Here we shall set down some of the methodologies of systems biology as we observe them. As a conceptual context coming from practitioners of systems biology, this may then serve for the further development of the philosophy of this science.

4.1. The goals of systems biology

A discussion of what is or should be the methodology of systems biology requires us to be explicit about our goals in systems biology. The main one, of course, is to understand more general principles underlying the behaviour and mechanistic workings of the complete biological systems that sustain life. After all, and as we discussed above, systems biology should be a science and not just a technology for analysing special cases. Systems biology should discover new scientific laws, which may relate as much to physical–chemical, organizational and fitness aspects as to biochemical principles. With respect to this aim, mathematics should not take the form of modeling but rather constitute a way of codifying proposed or verified laws or principles. A case in point is the connectivity law of metabolic control analysis (see Fell, 1996; Heinrich & Schuster, 1996; Kell & Westerhoff, 1986; Westerhoff & van Dam, 1987), which can be most strictly formulated after defining a new property (i.e. the elasticity, see above) in mathematical terms.

A second aim then is the ability to understand the inner workings of particular living systems. Ultimately this is best done by having a computational or mathematical model of the system in terms of its components and the quantitative nature of the interactions between them. Such a model could be the result of 'simulation' and 'fitting', the model being adjusted in terms of its structure and/or its parameter values, until it describes the observed system behaviour. That description may then constitute understanding. Such a description corresponds to a mechanistic explanation but now in the systems sense.

However, as in other kinds of modelling (Corne et al., 1999; Kell & Knowles, 2006) we want more: A third aim derives from the ability to make predictions about the possible future behaviour of the system on the basis of changes we might make to our models. This creates possibilities of further testing the quality of the model, which is the third aim of modelling. Using a model to make such predictions forbids its further adjustment whilst calculating the prediction; no fitting should be involved at such a stage. The same is true in machine learning (Duda et al., 2001; Hastie et al., 2001; Rowland, 2003). A related, fourth aim of modelling is the use of the model for technological or therapeutic purposes.

The fifth or ultimate aim of systems biology combines the above; it is the aim of accomplishing the mission of the life sciences and understand living systems in molecular terms, thereby opening such 'applied' avenues as prognosis, diagnosis, preventive medicine and lifestyle adjustment, therapy, drug design and biotechnology.

Here we have addressed the understanding of biological systems more than their explanation in an evolutionary context. Where we addressed explanation this is in terms of the direct causal mechanisms rather than those that derive from divergence and selection for fitness or stability or observability. After all, biological systems live in the absence of evolution. Our discussion has

also refrained from discriminating explictly between the two chief strategies for scientific understanding, i.e. by unification through subsumption to laws and understanding in terms of causation through mechanisms.

4.2. Systems biology: What it is

From the above aims and from the background of the limitations of molecular biology and functional genomics, one may surmise which activities are necessary for a successful systems biology. Many of the tools and techniques of functional genomics are in place as are the techniques from molecular biology and biochemistry. In view of the complexity of the subject matter, and because a focus on parts is ultimately not advised, our present strategy is to focus on a single system of life that is relatively autonomous. Ultimately this should result in a complete living organism being the object of study, and as scientific data and knowledge become distributed and available to all via the Internet this is increasingly possible in a coherent manner. At first these are likely to be unicellular microorganisms, or relatively autonomous subsystems thereof. The mathematical tools will be discussed in more detail below.

Figure 2 therefore shows some of the elements of the systems biology agenda (Kell, 2006). It gives a certain primacy to the system of interest as a circle in the centre. However, while specifics of methods will vary between organisms and

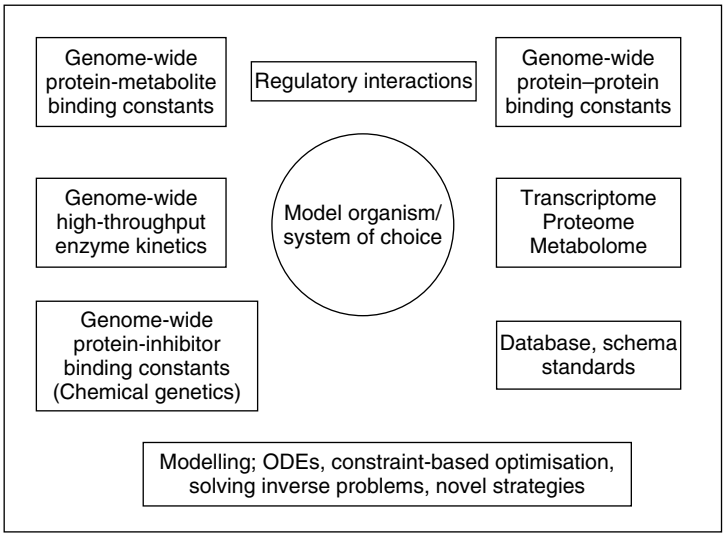

Figure 2 Some elements of the systems biology agenda.

These are purposely not interconnected in this figure for reasons of clarity.

systems (e.g. the optimal extraction method for the transcriptome of *Streptomyces coelicolor* – an organism with an unusually high GC content – differs substantially from that for the transcriptome of other organisms), we shall more or less ignore these specifics and here concentrate on generic issues and methodologies.

4.3. The spiral of knowledge

We maintain that for systems biology as well as for science generally, scientific thinking should consist of an interplay between (i) the mental worlds of knowledge and ideas and (ii) the physical world of observations and sense-data. Figure 3 sketches a straightforward view of the relationships between the two worlds, which is usually described as a cyclic interplay between experimental observation and theory, with induction on the basis of experimental observations leading to new, more acute experiments testing the hypotheses. The new experiments should then lead to a further adjustment of the intellectual world view and good hypotheses that derive therefrom. We note then that functional genomics without the systems biology dimension might remain in such a cycle of data collection, pattern recognition and the generation of ad hoc empirical 'laws' and hypotheses describing those data phenomenologically. The application of

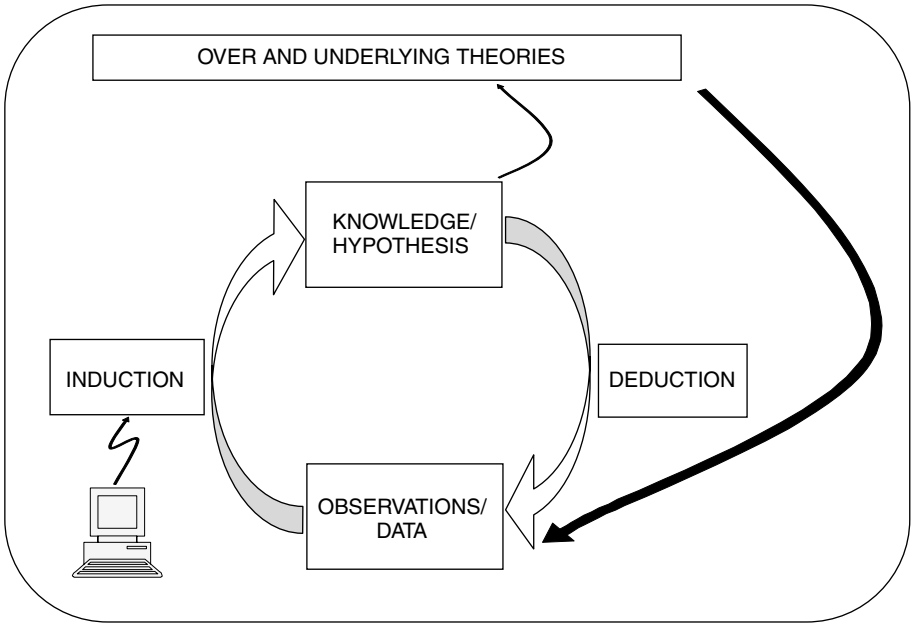

Figure 3 An iterative interplay between the world of ideas and the world of data as the hallmark of both science and systems biology.

systems biology in addition to functional genomics should lead to a progression of insight that is also outside the range covered by the primary dataset. The developing insight is effectively a third dimension, which is one of the aspects that systems biology may help add to functional genomics.

An example would be the observation in a large number of datasets that mRNA for a protein A always goes up or down together with that of protein B. This would lead to the empirical law that proteins A and B always behave similarly. This empirical law would reside on the same conceptual plane as the primary data set and would therefore fit into the cycle picture of Fig. 3. Here the broad aim of functional genomics could be seen to have been satisfied, and experimentation could stop. However, systems biology would search further for the cellular control and regulation hierarchy to find that the two corresponding genes are regulated by the same transcription factor; it would then search for interactions responsible for the correlation. Not only would this explain the observed correlation of mRNA-A and mRNA-B, it would also predict exceptions to these correlations, e.g. when a second transcription-factor footprint would map to gene A but not to gene B. In this way understanding will slowly but steadily grow outside the primary data set and elucidate more and more of cell biology, hence add a dimension of understanding.

We therefore recognize that systems biology may be among the sciences that is better described by a spiral of knowledge rather than a cycle (cf. Fig. 4). A further addition to the traditional vision is that of a box with overlying and underlying theories, with a deductive arrow stemming from that (cf. Figs. 3 and 4). Indeed, any law or hypothesis of systems biology should be consistent with underlying physical–chemical principles and in good systems biology any such hypothesis should therewith also be deduced in part from those underlying principles (this may seem a superfluous remark but we have seen systems biology-type theories that were inconsistent with the second law of thermodynamics and principles of electric fields).

4.3.1. Systems biology: The inductive versus the deductive mode

The recent developments in postgenomics have caused the empirical branch of systems biology, which is closest to functional genomics and stems from the developments in molecular biology (Westerhoff & Palsson, 2004), to develop most strongly. This branch emphasizes the observation component, i.e. the measurement of the dynamic variables. It then establishes patterns in the observed dynamic responses of the system to perturbations, whereby it uses mathematics for the analysis of multidimensional systems. This functional genomics activity tends towards systems biology because it accommodates the feature that the various molecules in the living cell vary coordinately in concentration. Often it is not yet the science of systems biology because it sticks to the observation of the correlations, without necessarily understanding their basis or whether they are

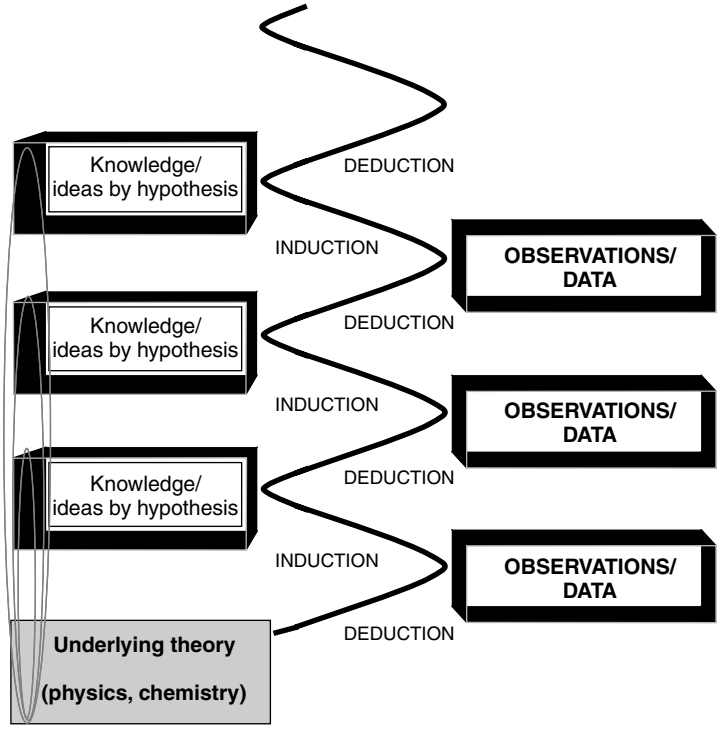

Figure 4 The advancement of Science and of Systems biology as a spiral.

Since the hypotheses are (hopefully) not the same at each turn of the cycle of Fig. 3, one may also or better view the iterative interplay between the elements of Fig. 3 in terms of a spiral.

in an explicit sense causal. This is not to say that this activity is not extremely useful, however, since observations of correlations between the transcriptome of tumours and their response to chemotherapy may help therapy tremendously, long before any mechanistic basis for understanding (and one might comment that this is widely true in medicine).

Functional genomics does become part of the science of systems biology when it makes the step of induction of Fig. 3. In practice, this has not yet happened very often. It seems important to redress the balance by transforming this empiricism into a principled hypothesis-generating arc that leads from data to knowledge. One way in which this can be done is to map the mRNA concentrations that vary coordinately onto the known regulatory maps of cell biology. Perhaps this leads to the recognition of coherent regulation of a pathway, or of a limited number of super-regulators. Either result would lead to a hypothesis which could then be tested further.

The deductive mode of reasoning is a classical obsession of biology, and remains entirely relevant. In the present context, it ranges from branches of

mathematical biology and metabolic control analysis which have been deduced from underlying principles, to proposed flux patterns (Reed & Palsson, 2003), or distributions of control (e.g. Hornberg et al., 2005).

By contrast, much of postgenomics and systems biology, in which often we lack reasons or sufficient background knowledge that might lead us to realistically plausible hypotheses, has been data-driven, with a good hypothesis being the result, not the starting point, of the initial investigation. This brings with it a requirement for a different kind of experimental design, in which rather than seeking to hold everything constant except one parameter we seek to vary conditions as much as possible (but in a controlled manner!) to produce a 'training set' of data to establish rules that are likely to generalize well to apply to examples not previously encountered (Kell & King, 2000). This entirely different way of thinking also discriminates the methods of classical statistics (that start with a model and test the goodness of fit of data to that model) from those of machine learning (that start with data and determine the model that best fits those data) (Breiman, 2001).

The chief element of this integrated view of the relation between ideas and data is the recognition that induction is not simply the reverse of deduction (Carnap, 1966; Kell & Welch, 1991). Deductive reasoning starts with an axiom or set of axioms (i.e., a mental construct, the world of ideas, such as 'all swans are white') and a hypothesis such as 'Alice is a swan' that together allow one to deduce with logical certainty that provided Alice is a swan one may make an observation in the expectation that Alice will be found to be white and the data found to be consistent with the hypothesis. Alternatively if Alice is found to be black then either Alice is not a swan or the axiom should be modified (axioms are by definition true). This hypothetico-deductive framework, in which hypotheses can be falsified by data but not proved true, was the focus of Karl Popper's agenda to demarcate 'science' from 'pseudo-science' (Medawar, 1982; Popper, 1992), although one must remark that in the real world some favoured hypotheses can survive in the face of any number of inconvenient facts (Gilbert & Mulkay, 1984; Kell, 1988; Kuhn, 1996).

The inductive mode of reasoning generalizes from patterns observed in a number of actual cases, and thus goes from the world of data to the world of ideas: If Alice is a Swan and is white, Bob is a swan and is white, and George is a swan and is white, an induction might be that 'all swans are white'. Now it has been known since the time of Hume that such induction is logically insecure, in the sense that a single black swan shows it, and that the fact that the sun has risen every morning throughout one's life does not mean it will provably do so tomorrow. However, the existence of black swans is no less harmful to the hypothesis on which the deduction is based that all swans are white than it is to the same view arrived at inductively, and it is not at all clear why induction should in fact be so disfavoured.

The systematic genome sequencing programmes did not set out with any specific hypotheses, save that the provision of such data might be of value (Kell & Oliver, 2004), and Sulston has stressed the importance of hypothesis-free measurements at appropriate stages in the growth of a science (Sulston & Ferry, 2002). Equally, the development of technology is also free of specific hypotheses (again save that their availability would be of scientific value), and it is hard to imagine working in a modern laboratory without techniques (cloning, sequencing, PCR, mass spectrometry, etc.) that have only been available for a comparatively short time (and many of which secured Nobel prizes for their developers). Equally, we see that many measurements, especially in postgenomics (Kell & King, 2000), are designed to be data-driven rather than hypothesis-driven (hypothesis-dependent). Thus in systems biology, science advances by an iterative and spiralling interplay between deductive and inductive reasoning, with a substantial amount of technology development also involved.

Our description of the (preferred) development of systems biology as a spiral, should not be taken to imply that we think of this as unique to systems biology. The development of many other natural sciences may be and have been described in similar terms. They can easily be represented as 'the cycle of knowledge' (Fig. 3).

It should also be mentioned that in many presentations of the novelty of systems biology to audiences of biologists, physicists and chemists, the cycle of knowledge is presented as something that can now finally be brought into effect in biology. This has reasons. First, in biology the experimental activities have become so complex and extensive, and demand such extensive experimental expertise, that the corresponding scientists have had little opportunity to engage in the complete cycle of knowledge. Second, molecular cell biology has long been incomplete in the sense that at any moment an as yet unknown molecule could turn up and explain experimental phenomena without having implications of the theories being tested or examined. For instance, when a hypothetical regulatory effect proposed by a theory is tested by an experiment, an additional, parallel effect would most often turn up, incapacitating the experimental testing of the theory. With functional genomics, it has become possible to have a complete inventory of virtually all relevant molecules, removing this limitation to the testing of theories. Third, in the case of systems biology, the complexity is often so great that the experimental and theoretical parts of the cycle cannot be within the expertise of the same individual. Therewith the cycle of knowledge is also relevant to indicate the roles various individuals in a project have with respect to each other.

4.3.2. Systems biology: The top-down/analytic versus the bottom-up/ synthetic strategies

Strategies and methodologies for systems biology come in a number of flavours, often discriminated as top-down and bottom-up, but also potentially including

middle-out (e.g. Brenner, 2001; Noble, 2003). While the true understanding of complex living systems and/or their subsystems will likely involve the judicious and iterative blending of each, it is convenient to use this distinction as a means of discriminating the necessary methodologies.

Analytical or top-down systems biology tends to start from the system as a whole. In a way it comes from the direction of holism and moves towards molecular mechanism. Either from empirical relations between genome-wide patterns of gene expression, or by calculating properties of genome-wide networks, it induces or proposes the occurrence of more general principles, such as the feature that metabolic networks correspond to small world, scale-free networks (Barabási & Oltvai, 2004; Wagner & Fel, 2001) and that genetic networks abound in certain regulatory motifs (Itzkovitz & Alon, 2005; Milo et al., 2002; Yeger-Lotem et al., 2004). These views may then be tested.

In the leaner, 'Synthetic' or bottom-up branch of systems biology, one typically starts with a qualitative ('structural') and often simple model of molecules interacting with each other in networks, then seeks to determine what system properties might emerge from the nonlinear interactions. By then parameterizing the equations that describe these interactions and inserting parameter values that correspond to actual subsystems, more or less realistic predictions of system properties are achieved. When the predictions are accurate, the proposed mechanisms of emergence of the functional properties are considered to have become more likely. This method is reductionist in that it prefers to deal with simple parts of the true system but not so simple as to lose important aspects of the interactions and the emergence of interesting functional properties. 'Bottom-up' methods start with purified entities (e.g. proteins) that allow the measurement of the parameters, while 'top-down' methods seek to infer their values via 'reverse engineering' of the parameters values through fitting of the calculated system behavior to experimentally observed system behaviour.

4.3.3. The bottom-up approach to systems biology

Our own prejudices – given a historical focus more on metabolic than signalling systems (Kell et al., 1989; Kell & Westerhoff, 1986; Mendes et al., 1996; Pritchard & Kell, 2002; Raamsdonk et al., 2001; Teusink et al., 2000; Westerhoff & Kell, 1987; Westerhoff & Kell, 1988; Westerhoff & Kell, 1996; Westerhoff et al., 1991), and on unicellular organisms rather than the more obviously (cf. Davey & Kell, 1996; Kell et al., 1991) differentiated 'higher' organisms – leads us to concentrate more on the 'bottom-up' approach (Fig. 5), embodied in the 'silicon-cell' concept (Westerhoff, 2001): if we can measure all of the 'local' properties of individual players in a complex system, including their interactions, we can bolt the system together and whatever new properties may emerge will indeed emerge and produce the 'whole system' properties that can indeed be compared with those of the intact system. The apotheosis of this approach to

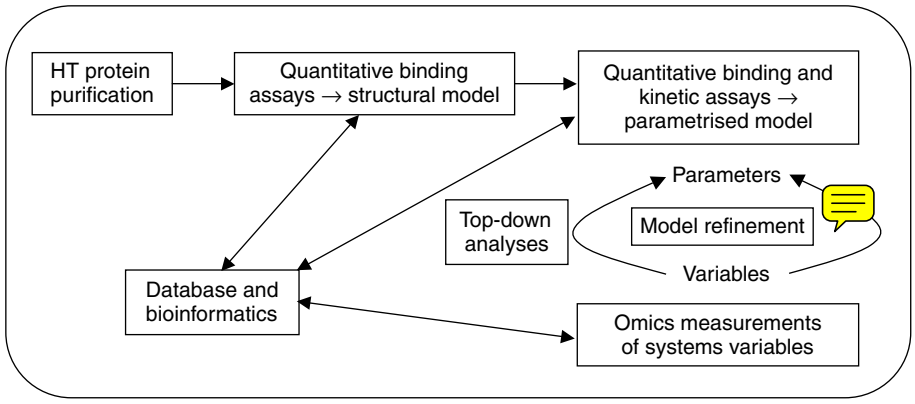

Figure 5 A largely 'bottom-up' strategy for systems biology.

date is the demonstration that the operation of yeast glycolysis under particular conditions can indeed be rather well predicted on the basis of the 'properties' of the isolated enzymes which participate in the overall process (Teusink et al., 2000) (and see (Pritchard & Kell, 2002)). It takes its strongest form when the interactive properties of all the relevant components of the system are put into a precise mathematical model, that is a computer replica ('silicon cell', see below) of the actual system; and if the system behaviour is then calculated successfully.

Occasionally it is argued that such a silicon-cell replica of an actual living cell would be completely reductionistic and therewith incapable to deal with the systems biology of the living cell. This is incorrect. Save for vital force influences, and given an initial physiological condition (cf. below), all there is in the living cell, at least in one way of looking at it, is a large number of molecules and all their interactions. Therewith, all that matters is the components and the relational properties of those molecules. If molecules and interactions (in their spatial context) are precisely reproduced in a computer program, then all system behaviour should emerge. The crux resides in the live interaction between the molecules both in the cell and in the computer program. Here one type of macromolecule carries out a process for a little while, by which it changes its environment in terms of a few, nameable properties such as the concentration of micromolecules like ATP, whilst leaving the rest of its environment unaltered (see below). The change in environment leads to a change in behaviour of other types of macromolecules in the same environment in the same cell (e.g. other enzymes in the same metabolic pathway). The altered behaviour of the latter molecules will again change the environment of the first macromolecule and therewith the behaviour of the former. In this way the activity of the first molecule depends on its own properties through the dynamic activities of the other molecules. Loosely formulated, it is the resonance with other molecules that

determines much of the behaviour of each individual molecule. In biology, this part of the molecule's behaviour often leads to important function. An example of the molecular behaviour that only originates in the dynamic interactions with the other molecules, is found with the molecules that are 'responsible for' the cell cycle. None of these would have a cyclic activity in the absence of the others, and this collective cycling is assumed to be the only biological function of these molecules.

The ultimate silicon-cell strategy completely recovers the emergence of functional behaviour of molecules from this resonating with the other molecules. A completely reductionistic approach would look only at the behaviour of the individual molecules, perhaps in an environment that is a frozen representation of the molecules' environment in the living organism. It then sees the behaviour of the living organism as the sum of these molecular behaviours, and thereby misses the extra molecular behaviour that stems from the cycle of interactions running through the other molecules. It would not comprehend the cell cycle, as it would perhaps observe but not explain the cycling.

An important issue is whether the silicon cell requires only molecular knowledge or also systems knowledge to start from. For sure, it does not require systems knowledge of the resonating type (cf. above). On the other hand, the systems of interest are nonlinear and the response of the molecules to the changes in their immediate environment do depend on the average state around which these changes occur, such as intracellular pH and ionic strength. The latter are indeed established by the system as a whole, and in this sense systems properties that correspond to the static physiological state do enter the silicon-cell models. These properties are static in the sense that they could be determined by taking a photograph (Kell and Mendes, 2000), or when they are time dependent, by a movie of the system around the macromolecule of interest. These properties are essentially parameters for the functioning of the interacting macromolecules, whereas the properties that create emergent properties are dynamic variables (cf. below).

As in fundamental physics, there could be cases where it is not really possible to consider macromolecules separately from their molecular environments. In these cases, their complete environment is codetermined by the dynamic behaviour of the macromolecules of interest. Then also, that entire environment consists of variables that are influenced by the macromolecules under study. This might (but would not have to) happen with regards to amino-acid residues in the system of the surrounding amino acids in a protein, or in MAP kinase cascades when all the kinases and phosphatases form a supercomplex, a scaffold.

The silicon-cell approach assumes that there is substantial possibility to consider macromolecules separately from their environments. In cases where parts of that immediate environment is not separable, that part needs to be taken together with the macromolecule. This then still does not incapacitate the silicon-cell

approach. If the inseparability is so massive that effectively the entire living cell has to be treated as a single macromolecule, the silicon cell approach does become impractical.

This issue has been alluded to in Boogerd et al. (2005). In the philosophical sense, they have defined the generation of new properties in those systems where macromolecules can be considered as separable from their physical–chemical environment as weak emergence. The cases where macromolecules are not separable from their environment would lead to strong emergence. We would here suggest that it will be possible to make all essential properties of living organisms emerge from silicon-cell-type models. This then implies that all functional properties of living systems come from weak emergence. We base this conjecture on the experience that free-energy transduction, gene expression, cell cycling and developmental biology can be generated by such models (cf. www.siliconcell.net). However, it is a conjecture at present; although these functional properties can be calculated, it has not been verified by experimental testing whether the models generate the functional properties in a quantitatively correct way and from the actual kinetic properties of the constituent macromolecules. And then, there are cases where function arises, where such calculations have not yet been possible, such as in the cases of epigenetic regulation of gene expression.

4.3.4. Parameters and variables and who controls whom

An important distinction to be made in systems biology (and not only there) is between parameters and variables. Parameters are elements set to fixed values by the system itself or controlled externally by the experimenter, while variables are those elements that change during the course of an experiment. (Note that the elapsed time, though in fact a variable, is normally considered an honorary parameter.) In an isolated metabolic system in which protein synthesis and degradation are not occurring, the parameters are then the concentrations, and especially the kinetic and binding constants, of the enzymes involved, as well as the 'fixed' concentration of 'external' substrates. The variables are then the time-dependent concentrations of the intermediary metabolites and the flux(es) through the pathway or network of interest. Two facts are to be noted. First, only parameters can control variables; and variables cannot control other variables. Parameters are controlled neither by other parameters nor by variables. Secondly, normally it is variables that are measured experimentally, as such measurements of changes are easier – and this statement includes all the 'omics' ('expression profiling') methods such as transcriptomics, proteomics and metabolomics. Given these facts, it is seen that there has therefore been a very great dearth of systematic measurements of the properties that we actually wish to measure, viz. the binding and kinetic constants of individual proteins (and other molecules). Such measurements were commonplace in the 1960s and early 1970s (a large

number of papers in the journal *Biochemistry* at that time were entitled 'purification and properties of some enzyme'), and we need these times to return to biology, with concomitant modernization of the way in which and the scale at which the experiments are done. Indeed, in an account of what needs to be done by bottom up systems biology, one finds many 'old-fashioned' looking terms (cf. Table 1).

4.3.5. *Strategies for determining binding and kinetic constants for individual proteins*

In the spirit of Mrs Beeton (Beeton, 2000), 'first get your protein'. While these will still require purification, often via dual affinity tags, they will normally

Table 1 Some methodologies of significance for 'bottom-up' systems biology

Stages	Methodologies	Comments	Selected references
'First get your protein'	Cloning, expression and purification	Choice of hosts and vectors, tags, growth media, glycosylation and refolding	
Qualitative binding assays	Mass spectrometry and FTIR	Allows production of a structural model. The binding of some elements may depend on that of others.	(Muckenschnabel et al., 2004; Wharton, 2000; Zehender et al., 2004)
Quantitative binding assays	Mass spectrometry	High-resolution methods such as FTICR are useful	(Last & Robinson, 1999)
High-throughput kinetic methods	Optical, mass spectrometry and calorimetry		(Shen et al., 2004; Ward & Holdgate, 2001)
Omics measurements	Microarrays and mass spectrometry		(Aebersold & Mann, 2003; Goodacre et al., 2004; Schena, 2000)
Bottom-up model	ODE modelling		(Mendes & Kell, 1998)

be prepared by recombinant means. We shall not deal here in detail on these methods, save to note that the systematic production of nominally *all* the proteins of baker's yeast (*S. cerevisiae*) has been performed by Snyder and colleagues (e.g. Phizicky et al., 2003; Zhu et al., 2001) and in this sense the industrialization of such processes has begun (see also, e.g. for *C. elegans* – http://sgce.cbse.uab.edu/). It is also worth pointing out that even in well-established recombinant hosts there is a nonlinear interplay between the specifics of the recombinant vector, the exact host strain and the growth and production media used to induce the synthesis of the target protein of interest in a form that allows successful purification and refolding.

The next stage is represented by qualitative binding assay, by which we seek the 'structural model' that describes the players including substrates, products and effectors of enzymatic reactions, protein–protein and protein–nucleic interactions and so on (see Fig. 5).

4.4. The special role of mathematics in systems biology: Calculating emergence

As do most commentators (e.g. Hood, 2003; Ideker et al., 2001; Kitano, 2002; Naylor, 2004/2005), we (Kell, 2004; Kell, 2005; Kell, 2006; Kell & Knowles, 2006; Westerhoff & Palsson, 2004) consider systems biology to involve an interplay between theory, computation/modelling and experimental activities. This interplay is strongly catalysed by the development of new technologies, and in fact it is these developments more than anything else that has accelerated the subject (Hood, 2003). It should be noted that Fig. 6 differs rather significantly from Fig. 3, which we presented as our standard paradigm for scientific activity. Indeed, we should like to suggest that in systems biology as in other systems sciences, the role of mathematics is more fundamental than it is in sciences that deal with single entities of much lower inherent complexity.

Of course, mathematics helps the analyses of the rather complex datasets in helping to establish correlations, which then feed into the inductive mode of Fig. 3. It helps ordering the data, then remaining in the empirical box of Fig. 3. It also helps formulate the hypotheses and theories inside the box theory of Fig. 3. And it may help deduce experimental implications from the theories, helping the deductive process depicted in Fig. 3. The reasons for modeling are numerous, and covered elsewhere (Kell & Knowles, 2006; Klipp et al., 2005), and include testing whether the model is accurate, in the sense that it reflects, or can be made to reflect, known experimental facts, analysing the model to understand which parts of the system contribute most to some desired properties of interest, hypothesis testing, allowing one to analyse the effects of manipulating experimental conditions in the model without having to perform complex and costly experiments, and seeing what changes in the model would improve the

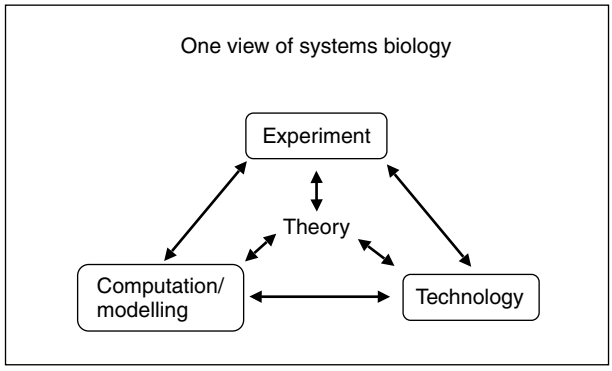

Figure 6 Systems biology as an iterative interplay between theory, experiment and technology development and modelling.

consistency of its behaviour with experimental observations. While these roles of mathematics may be stronger in systems biology than in other sciences, they are not qualitatively different.

The special role of mathematics (which we take to include numerical computation) in systems biology derives from the following. It is an aim of systems biology to understand how properties emerge in the interactions of components of systems. The emergence of these new properties should be completely determined by all those interactive properties. If the interaction properties of the components are correctly known on the basis of experiments with the individual molecule species, then emergence of the new properties in a precise computer model is inescapable. The very emergence is thus not in this direct sense subject to experimental testing. In this aspect systems biology is not subject to experimental testing either. It may be subject to computational testing, however.

In molecular biology similar situations may arise. The properties of a molecule are proposed to have an effect on its behaviour, such as that the adjacency of two glutamate residues in a protein are responsible for the binding of calcium. Usually in molecular biology no time nor effort is wasted in calculating whether indeed in principle the adjacency of the two glutamate residues could enhance calcium binding; this is considered 'obvious' (actually, it may not be quite obvious; protein dynamics calculations should perhaps be carried out; but in view of the many nonlinear interactions involved, this is akin to invoking systems biology). In systems biology it is more often not trivial to see whether a proposed mechanism for emergence could account for the emerging property, even independent of whether the proposed interactive properties are real experimentally. It involves a computational experiment to check if indeed the proposed interactions could generate the emergent behaviour. This is so because the interactions are so complex that an immediate intuitive prediction is impossible, and because the

emergence depends on the particular magnitude of the parameter values, i.e. on the particular condition the system is in. (We note, though, that in a sense, such questions about protein engineering are not quantitative, since changing one or both of a pair of adjacent glutamates to alanine may perfectly well change the structure and dynamics of the enzyme irrespective of any effect on their ability to bind calcium.) It is of course well known that even simple systems can exhibit very complex dynamics (Abraham & Shaw, 1992; May, 1976). Accordingly, computation here plays something of the role of experimentation in other sciences. The hypothesis that an experimentally established set of interactions is responsible for certain emergent behaviour in the system needs to be tested by performing calculations.

Although this situation is new to much of the life sciences and was not made very explicit in the original philosophies of physics (Carnap, 1966), it is standard to present-day physics and chemistry. In particle physics and in statistical thermodynamics, certain properties may be known experimentally. The question is then asked whether those properties may be responsible for certain observed behaviour, and the answer is obtained solely by numerical experimentation.

We recently carried out this type of numerical experimental systems biology when proposing that the compound acetaldehyde might be 'responsible' for the synchronization of glycolytic oscillations between individual yeast cells (Reijenga et al., 2005a). Putting in the actual structure of the network in so far as we could, we calculated that the synchronization should indeed occur. More recently, we posed the hypothesis that the glycolytic oscillations in yeast are not controlled at a single step such as the proposed pace-maker enzyme phosphofructokinase, but at many points in the network at the same time. Again numerical experiments based on what was already known experimentally about the interaction and networking in the system, served to verify the hypothesis in the numerical sense (Reijenga et al., 2005b).

We should like to emphasize that in no way do we wish to detract from the importance of experimental work for systems biology. If anything, experimentation is more important to systems biology than to molecular biology, in view of the strong dependence of what actually happens on the precise parameter values. It is just that mathematics is also more important to systems biology than it is to molecular biology.

4.4.1. Precision, silicon cells and the calculation of emergence

The calculations we referred to here are often deductive in the sense that they start from a hypothesis and calculate whether indeed the proposed mechanisms of emergence deliver the proposed emergent property. However, calculations in the sense of numerical experiments can also be used to induce general properties.

Indeed, this was involved in the origin of one of the more distinctive laws of systems biology, i.e. the summation theorem as discovered by Jim Burns and the late Henrik Kacser (Kacser, personal communication).

The emergence of properties from nonlinear systems depends on the values of the parameters. The consequence has long been overlooked by theoretical biologists and biologically inspired physicists. The latter supposed that it was good enough to show that some, phenomenological model of the biological system could produce the emergent property of interest. In this manner, Turing modelled developmental biology (in a way that is now known to be wrong, even though parts of the self-organization mechanisms may still act), and Nicolis and Prigogine modelled glycolytic oscillations in yeast. They did find that in such a phenomenological model (with oversimplified and in fact unrealistic rate equations and rather arbitrarily chosen parameter values) the emergent phenomena occurred. For different rate equations or different parameter values, the emergent property did not emerge from the calculations. Hence, to verify whether a proposed systems biology mechanism is indeed responsible for an observed emergent property, the model must be precise in terms of its structure and parameter values. Until recently the handicap was of course that such precise parameter values were not available. (Consequently, the above should not be taken to question the importance of this earlier work in biological physics and theoretical biology.)

With the advance of experimental techniques and thanks to the effort of many scientists, it is now becoming possible to make the required precise models. We refer to these precise models as 'computer replicas' of the real network of interactions or 'silicon cells' (Westerhoff, 2001). In a sense, the silicon cell strategy is entirely reductionist, yet at the same time upwardly compatible with holism (Snoep & Westerhoff, 2005). All the molecules known to act in a network are represented by a computer replica. At present this most often takes the form of a rate equation and a reaction equation for each enzyme. The rate equations, i.e. the reaction equations as well as the values of the parameters therein, should have been established experimentally (here we recognize the irreducibility discussed above) and are all inserted into the computer replica of the network. All the computer then does is let the replica behave through the integration of the equations in time. Emergent properties, if any, should then show up in the computer calculations (*modulo* the statistical error in the measurements).

In this manner, ordinary and partial differential equations may be used to calculate life, i.e. to produce a silicon cell that will display the main properties of the real cell, inclusive of the emergent properties. The implications are unprecedented for the sciences: If there is any place in the natural world where qualitatively new properties emerge, this is life.

In terms of philosophy, we are becoming iconoclastic here however. Emergent properties are sometimes defined as the properties that are irreducible. If properties can be calculated then by some kinds of definition they are not emergent. We consider this definition inappropriate, and it may stem from an oversight of the distinction between linear and nonlinear calculations. Properties that can be calculated from a linear superposition of properties of the components of a system (such as their total mass) should indeed not be called emergent. The important distinction comes when qualitatively new properties can be calculated in systems with essential nonlinear interactions. Only then are the properties new, they were not present in the components, and should indeed be said to 'emerge' (Solé & Goodwin, 2000, *pace* Boogerd et al., 2005).

We here make the challenging statement that life is calculable and can therefore be captured in a computer model. Within 10 or 20 years a silicon cell will have been constructed that accurately describes the main elements and behaviour of a living cell, and therefore can be rightfully considered a replica of the cell. Of course there are some exceptions with respect to a straightforward calculation of all aspects of life. These include deterministic chaos, systems that are extremely heterogeneous, and life beyond its simplest form already present in unicellular microorganisms. This said, a Digital Human, both generated and available in silico at a suitably coarse-grained level, will be a fantastic boon for both academic researchers and the Pharmaceutical industry alike; for the latter it may be expected to decrease substantially the present enormous attrition rates of candidate drugs. The issue of biological evolution too is much more important than suggested by our virtual lack of reference here. However, we have decided here to focus on life as it is at a certain moment in evolutionary history, not on how it came about in the sense of evolution. We think that the explanation of life as such is already a significant and challenging problem that requires systems biology for good answers. Perhaps with this treatise, and certainly with the entire book, we hope to have attracted Philosophers of science to a rapidly developing biology which may well be the place where things are happening in philosophy right now.

ACKNOWLEDGEMENTS

DBK thanks the BBSRC for financial support, and for keeping alive nonhypothesis-dependent science during the dark days and now, and Steve Oliver and Ross King for many interesting discussions. HVW thanks Fred Boogerd and Frank Bruggeman and many participants in the symposium for many thoughts and much food for thought, before, during and hopefully after the event. He is also grateful to the BBSRC, NWO, EU-FP6 and IOP-Genomics for various modes of support.

REFERENCES

Abraham RH & Shaw CD. *Dynamics: The Geometry of Behaviour.* Addison Wesley, Redwood City, CA, 1992.

Aebersold R & Mann M. *Mass spectrometry-based proteomics.* Nature: 422, 198–207, 2003.

Anderson S, Bankier AT, Barrell BG, Debruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R & Young IG. *Sequence and organization of the human mitochondrial genome.* Nature: 290, 457–465, 1981.

Bäck T, Fogel DB & Michalewicz Z. *Handbook of evolutionary computation.* IOP Publishing/Oxford University Press, Oxford, 1997.

Barabási A-L & Oltvai ZN. *Network biology: understanding the cell's functional organization.* Nature Reviews Genetics: 5, 101–113, 2004.

Beadle GW & Tatum EL. *Genetic control of biochemical reactions in Neurospora.* Proceedings of the National Academy of Sciences: 17, 499–506, 1941.

Beeton I. *Mrs Beeton's Book of Household Management.* Oxford Paperbacks, Oxford, 2000.

Beynon RJ, Doherty MK, Pratt JM & Gaskell SJ. *Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides.* Nature Methods: 2, 587–589, 2005.

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, ColladoVides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B & Shao Y. *The complete genome sequence of Escherichia coli K-12.* Science: 277, 1453–1462, 1997.

Boogerd FC, Bruggeman FJ, Richardson RC, Stephan A & Westerhoff HV. *Emergence and its place in nature: A case study of biochemical networks.* Synthese 145: 131–164, 2005.

Breiman L. *Statistical modeling: The two cultures.* Statistical Science: 16, 199–215, 2001.

Brenner S. Discussion. In: Complexity in biological information processing (Eds.: Bock G & Goode J), Wiley, Chichester, 150–159, 2001.

Brent R. F*unctional genomics: Learning to think about gene expression data.* Current Biology: 9, R338–R341, 1999.

Carnap R. *Philosophical foundations of physics.* Basic Books, New York, 1966.

Corne D, Dorigo M & Glover F. *New ideas in optimization.* McGraw Hill, London, 1999.

Davey HM, Davey CL, Woodward AM, Edmonds AN, Lee AW & Kell DB. *Oscillatory, stochastic and chaotic growth rate fluctuations in permittistatically-controlled yeast cultures.* Biosystems: 39, 43–61, 1996.

Davey HM & Kell DB. *Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analysis.* Microbiological Reviews: 60, 641–696, 1996.

Duda RO, Hart PE & Stork DE. *Pattern classification*. 2nd ed. John Wiley, London, 2001.

Dunn WB, Bailey NJC & Johnson HE. *Measuring the metabolome: current analytical technologies.* Analyst: 130, 606–625, 2005.

Dunn WB & Ellis DI. *Metabolomics: current analytical platforms and methodologies.* Trends in Analytical Chemistry: 24, 285–294, 2005.

Fell DA. *Understanding the control of metabolism.* Portland Press, London, 1996.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM & et al. *Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.* Science: 269, 496–512, 1995.

Frauenfelder H & McMahon BH. *Relaxations and fluctuations in myoglobin.* Biosystems: 62, 3–8, 2001.

Fröhlich H & Kremer F. *Coherent Excitations in Biological Systems.* Springer, Berlin, 1983.

Gilbert GN & Mulkay M. *Opening Pandora's box : a sociological analysis of scientists' discourse.* Cambridge University Press, Cambridge, 1984.

Gleick J. *Chaos: making a new science.* Abacus, New York, 1988.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H & Oliver SG. *Life With 6000 Genes.* Science: 274, 546–567, 1996.

Goldbeter A, Gonze D, Houart G, Leloup JC, Halloy J & Dupont G. *From simple to complex oscillatory behaviour in metabolic and genetic control networks.* Chaos: 11, 247–260, 2001.

Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG & Kell DB. *Metabolomics by numbers: acquiring and understanding global metabolite data.* Trends in Biotechnology: 22, 245–252, 2004.

Hastie T, Tibshirani R & Friedman J. *The elements of statistical learning: data mining, inference and prediction.* Springer-Verlag, Berlin, 2001.

Heinrich R & Schuster S. *The regulation of cellular systems.* Chapman & Hall, New York, 1996.

Hill TL. *Free Energy Transduction in biology: Steady State Kinetic and Thermodynamic Formalism.* Academic Press, New York, 1977.

Hood L. *Systems biology: integrating technology, biology, and computation.* Mechanisms of Ageing and Development: 124, 9–16, 2003.

Hornberg JJ, Bruggeman FJ, Binder B, Geest CR, de Vaate AJ, Lankelma J, Heinrich R & Westerhoff HV. *Principles behind the multifarious control of signal transduction. ERK phosphorylation and kinase/phosphatase control.* FEBS Journal: 272, 244–258, 2005.

Ideker T, Galitski T & Hood L. *A new approach to decoding life: systems biology.* Annual Reviews of Genomics and Human Genetics: 2, 343–372, 2001.

Ihekwaba A, Broomhead DS, Grimley R, Benson N & Kell DB. *Sensitivity analysis of parameters controlling oscillatory signalling in the NF-kB pathway: the roles of IKK and IkBa.* Systems Biology: 1, 93–103, 2004.

Ihekwaba AEC, Broomhead DS, Grimley R, Benson N, White MRH & Kell DB. *Synergistic control of oscillations in the NF-kB signalling pathway.* IEE Proceedings Systems Biology: 152, 153–160, 2005.

Itzkovitz S & Alon U. *Subgraphs and network motifs in geometric networks.* Physical Review E; Statistical, Nonlinear, and Soft Matter Physics 71, 026117, 2005.

Keizer J. *Statistical Thermodynamics of Nonequilibrium Processes.* Springer, Berlin, 1987.

Kell DB. *Protonmotive energy-transducing systems: some physical principles and experimental approaches.* In: Bacterial Energy Transduction (Ed.: Anthony CJ), Academic Press, London, 429–490, 1988.

Kell DB. *Metabolomics and systems biology: making sense of the soup.* Current Opinions in Microbiology: 7, 296–307, 2004.

Kell DB. *Metabolomics, machine learning and modelling: towards an understanding of the language of cells.* Biochemical Society Transactions: 33, 520–524, 2005.

Kell DB. *Metabolomics, modelling and machine learning in systems biology: towards an understanding of the languages of cells.* The 2005 Theodor Bücher lecture. FEBS Journal: 273, 873–894, 2006.

Kell DB, Brown M, Davey HM, Dunn WB, Spasic I. & Oliver SG. *Metabolic footprinting and Systems Biology: the medium is the message.* Nature Reviews Microbiology: 3, 557–565, 2005.

Kell DB & King RD. *On the optimization of classes for the assignment of unidentified reading frames in functional genomics programmes: the need for machine learning.* Trends in Biotechnology: 18, 93–98, 2000.

Kell DB & Knowles JD. *The role of modeling in systems biology.* In: System modeling in cellular biology: from concepts to nuts and bolts (Eds.: Szallasi Z, Stelling J & Periwal V), MIT Press, Cambridge, 3–18, 2006.

Kell DB & Mendes P. *Snapshots of systems: metabolic control analysis and biotechnology in the post-genomic era.* In: Technological and Medical Implications of Metabolic Control Analysis (Eds.: Cornish-Bowden A & Cárdenas ML), (and see http://dbk.ch.umist.ac.uk/WhitePapers/mcabio.htm). Kluwer Academic Publishers, Dordrecht, 3–25, 2000.

Kell DB & Oliver SG. *Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era.* Bioessays: 26, 99–105, 2004.

Kell DB, Ryder HM, Kaprelyants AS & Westerhoff HV. *Quantifying heterogeneity: Flow cytometry of bacterial cultures.* Antonie van Leeuwenhoek: 60, 145–158, 1991.

Kell DB, van Dam K & Westerhoff HV. *Control analysis of microbial growth and productivity.* Society for General Microbiology Symposium: 44, 61–93, 1989.

Kell DB & Welch GR. *No turning back, Reductonism and Biological Complexity.* Times Higher Educational Supplement 9th August, 15, 1991.

Kell DB & Westerhoff HV. *Metabolic control theory: its role in microbiology and biotechnology.* FEMS Microbiology Reviews: 39, 305–320, 1986.

King RD, Whelan KE, Jones FM, Reiser PGK, Bryant CH, Muggleton SH, Kell DB & Oliver SG. *Functional genomic hypothesis generation and experimentation by a robot scientist.* Nature: 427, 247–252, 2004.

Kitano H. *Computational systems biology.* Nature: 420, 206–210, 2002.

Klipp E, Herwig R, Kowald A, Wierling C & Jehrach H. *Systems biology in Practice: Concepts, Implementation and Clinical Application.* Wiley/VCH, Berlin, 2005.

Kuhn TS. *The structure of scientific revolutions.* Chicago University Press, Chicago, 1996.

Lakatos I. *Philosophical Papers 1: The methodology of scientific research programmes.* Cambridge University Press, Cambridge, 1978.

Last AM & Robinson CV. *Protein folding and interactions revealed by mass spectrometry.* Current Opinion in Chemical biology: 3, 564–570, 1999.

Laughlin RB. *A different Universe: reinventing physics from the bottom down.* Basic Books, New York, 2005.

May RM. *Simple mathematical models with very complicated dynamics.* Nature: 261, 459–467, 1976.

Medawar P. *Pluto's republic.* Oxford University Press, Oxford, 1982.

Mendes P & Kell DB. *Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation.* Bioinformatics: 14, 869–883, 1998.

Mendes P, Kell DB & Westerhoff HV. *Why and when channeling can decrease pool size at constant net flux in a simple dynamic channel.* Biochimica Biophysica Acta: 1289, 175–186, 1996.

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

Muckenschnabel I, Falchetto R, Mayr LM & Filipuzzi I. *SpeedScreen: label-free liquid chromatography-mass spectrometry-based high-throughput screening for the discovery of orphan protein ligands.* Analytical Biochemistry: 324, 241–249, 2004.

Nagel E. *The Structure of Science: Problems in the Logic of Scientific Explanation.* Routledge, London, 1961.

Naylor S. *Systems biology, information, disease and drug discovery.* Drug Discovery World 6, 23–40, 2004/5.

Nelson DE, Ihekwaba AEC, Elliott M, Gibney CA, Foreman BE, Nelson G, See V, Horton CA, Spiller DG, Edwards SW, McDowell HP, Unitt JF, Sullivan E, Grimley R, Benson N, Broomhead DS, Kell DB & White MRH. *Oscillations in NF-kB signalling control the dynamics of target gene expression.* Science: 306, 704–708, 2004.

Nicolis G & Prigogine I. *Self-organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations.* Wiley, New York, 1977.

Noble D. *The future: putting Humpty-Dumpty together again.* Biochemical Society Transactions: 31, 156–158, 2003.

Peletier MA, Westerhoff HV & Kholodenko BN. *Control of spatially heterogeneous and time-varying cellular reaction networks: a new summation law.* Journal of Theoretical biology: 225, 477–487, 2003.

Phizicky E, Bastiaens PI, Zhu H, Snyder M & Fields S. *Protein analysis on a proteomic scale.* Nature: 422, 208–215, 2003.

Popelier PL & Joubert L. *The elusive atomic rationale for DNA base pair stability.* Journal of the American Chemical Society: 124, 8725–8729, 2002.

Popper KR. *Conjectures and refutations: the growth of scientific knowledge.* 5th ed. Routledge & Kegan Paul, London, 1992.

Primas H. *Chemistry, Quantum Mechanics and Reductionism.* Springer, Berlin, 1981.

Pritchard L & Kell DB. *Schemes of flux control in a model of Saccharomyces cerevisiae glycolysis.* European Journal of Biochemistry: 269, 3894–3904, 2002.

Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh M, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K & Oliver SG. *A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations.* Nature Biotechnology: 19, 45–50, 2001.

Reed JL & Palsson Bϕ. *Thirteen years of building constraint-based in silico models of Escherichia coli.* Journal of Bacteriology: 185, 2692–2699, 2003.

Reijenga KA, Bakker, BM, van der Weijden CC & Westerhoff HV. *Training of yeast cell dynamics.* FEBS Journal: 272, 1616–1624, 2005a.

Reijenga KA, van Megen YM, Kooi BW, Bakker BM, Snoep JL, van Verseveld HW & Westerhoff HV. *Yeast glycolytic oscillations that are not controlled by a single oscillophore: a new definition of oscillophore strength.* Journal of Theoretical biology: 232, 385–98, 2005b.

Richard P, Teusink B, Westerhoff HV & van Dam K. *Around the growth phase transition S. cerevisiae's make-up favours sustained oscillations of intracellular metabolites.* FEBS Letters: 318, 80–82, 1993.

Rosen R. *Life itself.* Columbia University Press, New York, 1991.

Rowland JJ. *Model selection methodology in supervised learning with evolutionary computation.* Biosystems: 72, 187–196, 2003.

Schena M. *Microarray biochip technology.* Eaton Publishing, Natick, MA, 2000.

Schrödinger E. *What is life?* Cambridge University Press, Cambridge, 1944.

Scrutton NS, Basran J & Sutcliffe MJ. *New insights into enzyme catalysis – Ground state tunnelling driven by protein dynamics.* European Journal of Biochemistry: 264, 666–671, 1999.

Shen Z, Go EP, Gamez A, Apon JV, Fokin V, Greig M, Ventura M, Crowell JE, Blixt O, Paulson JC, Stevens RC, Finn MG & Siuzdak G. *A mass spectrometry plate reader: monitoring enzyme activity and inhibition with a Desorption/Ionization on Silicon (DIOS) platform.* ChemBioChem: 5, 921–927, 2004.

Snoep JL, van der Weijden CC, Andersen HW, Westerhoff HV & Jensen PR. *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.* European Journal of Biochemistry: 269, 1662–1669, 2002.

Snoep JL & Westerhoff HV. *Silicon cells.* In: Systems biology (Eds.: Alberghina L & Westerhoff HV), Springer, Berlin, 2005.

Solé R & Goodwin B. *Signs of life: how complexity pervades biology.* Basic Books, New York, 2000.

Sulston J & Ferry G. *The common thread: a story of science, politics, ethics and the human genome.* Bantam Press, London, 2002.

Sutcliffe MJ & Scrutton M. *Enzymology takes a quantum leap forward.* Philosophical Transactions of the Royal Society A: 358, 367–386, 2000.

Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, Walsh MC, Bakker BM, van Dam K, Westerhoff HV & Snoep JL. *Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry.* European Journal of Biochemistry: 267, 5313–5329, 2000.

Vaidyanathan S, Broadhurst DI, Kell DB & Goodacre R. *Explanatory optimisation of protein mass spectrometry via genetic search.* Analytical Chemistry: 75, 6679–6686, 2003.

Wagner A & Fel, DA. *The small world inside large metabolic networks.* Proceedings of the Royal Society B: 1803–1810, 2001.

Walter G, Bussow K, Cahill D, Lueking A & Lehrach H. *Protein arrays for gene expression and molecular interaction screening.* Current Opinion in Microbiology: 3, 298–302, 2000.

Ward WH & Holdgate GA. *Isothermal titration calorimetry in drug discovery.* Progress in Medicinal Chemistry: 38, 309–376, 2001.

Westerhoff HV. *The silicon cell, not dead but live!* Metabolic Engineering: 3, 207–210, 2001.

Westerhoff HV, Hellingwerf KJ & van Dam K. *Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate.* Proceedings of the National Academy of Sciences USA: 80, 305–309, 1983.

Westerhoff HV & Hofmeyr JH-S. What is Systems Biology? From genes to function and back. In: Systems biology (Eds.: Alberghina L & Westerhoff HV), Springer, Berlin, 2005.

Westerhoff HV & Kell DB. *Matrix method for determining the steps most rate-limiting to metabolic fluxes in biotechnological processes.* Biotechnology Bioengineering: 30, 101–107, 1987.

Westerhoff HV & Kell DB. *A control theoretical analysis of inhibitor titrations of metabolic channelling.* Comments on Molecular and Cellular Biophysics 5, 57–107, 1988.

Westerhoff HV & Kell DB. *What BioTechnologists knew all along . . . ?* Journal of Theoretical Biology: 182, 411–420, 1996.

Westerhoff HV, Koster JG, van Workum M & Rudd KE. *On the control of gene expression.* In: Control of Metabolic Processes (Ed.: Cornish-Bowden A), Plenum Press, New York, 399–412, 1990.

Westerhoff HV & Palsson BO. *The evolution of molecular biology into systems biology.* Nature Biotechnology: 22, 1249–1252, 2004.

Westerhoff HV & van Dam K. *Thermodynamics and control of biological free energy transduction.* Elsevier, Amsterdam, 1987.

Westerhoff HV, van Heeswijk W, Kahn D & Kell DB. *Quantitative approaches to the analysis of the control and regulation of microbial metabolism.* Antonie van Leeuwenhoek 60, 193–207, 1991.

Wharton CW. *Infrared spectroscopy of enzyme reaction intermediates.* Natural Products Reports: 17, 447–453, 2000.

Wolf J, Passarge J, Somsen OJ, Snoep JL, Heinrich R & Westerhoff HV. *Transduction of intracellular and intercellular dynamics in yeast glycolytic oscillations.* Biophysical Journal: J 78, 1145–1153, 2000.

Yeger-Lotem E, Sattath S, Kashtan N, Itzkovitz S, Milo R, Pinter RY, Alon U & Margalit H. *Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction.* Proceedings of the National Academy of Sciences USA: 101, 5934–5939, 2004.

Zehender H, Le Goff F, Lehmann N, Filipuzzi I & Mayr LM. *SpeedScreen: the "missing link" between genomics and lead discovery.* Journal of Biomolecular Screening: 9, 498–505, 2004.

Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M & Snyder M. *Global analysis of protein activities using proteome chips.* Science: 293, 2101–2105, 2001.