Part I
Biological Basis of Systems Biology

Systems Biology: Advances in Molecular Biology and Medicine, First Edition. Edited by Robert A. Meyers.

1
Systems Biology

Melanie Boerries[1], Roland Eils[2], and Hauke Busch[1]
[1]Freiburg Institute for Advanced Studies – LifeNet, School of Life Sciences, Albertstraße 19, 79104 Freiburg, Germany
[2]German Cancer Research Institute, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

Systems Biology: Advances in Molecular Biology and Medicine, First Edition. Edited by Robert A. Meyers.

Keywords

Systems biology
A new field of biology that studies the functional structure and dynamics of intercellular and intracellular networks with the help signal- and systems-oriented methods.

Synthetic biology
Studies life as networks of biological objects such as DNA proteins RNA and metabolites.

Network biology
Studies the static organization of life as networks made up of biological entities such as DNA proteins RNA or metabolites.

System
A set of interacting parts functioning as a whole and distinguishable from its surroundings by identifiable boundaries.

Systems theory
This denotes the cross-disciplinary investigation of the abstract organization of systems irrespective of their substance type or spatiotemporal scale of existence. The goal is the study of emerging properties that arise from the interconnectedness of the individual parts making up the system.

Robustness
The robustness of biological systems denotes the maintenance of specific system functionalities in the presence of fluctuations or change in environmental parameters.

Control
Control is defined as the response action taken by a system to counteract parameter changes to maintain system functions at a certain predefined level.

Modularity
A design concept of complex systems to integrate simpler self-contained functional building-blocks into the framework of one larger system.

Model
The concept of representing causal relationships from real systems in the language of mathematics.

Systems Biology is a new field of biology, which places the theoretical foundations of systems analysis of living matter into the context of modern high-throughput quantitative experimental data, mathematics, and *in silico* simulations. The aim is to analyze the organization and to gain engineering-control of metabolic and genetic pathways. The ultimate goal is to gain an ''holistic'' view of the complex workings of life. The need for a system level understanding of biology is reviewed in this chapter, and comments are provided on the current scientific progress in this field. The current and future directions of experimental design strategies and theoretical approaches are also highlighted.

1 Introduction

Systems Biology is a recently established field in life sciences that aims at promoting a global systems understanding of living matter through the integration of various scientific domains (see Refs [1–3] for special journal sections and Refs [4–7] for textbooks on the topic).

The considerate attention that Systems Biology receives is due to the fact that it currently causes a paradigmatic shift in many areas of biological research. Modern molecular biology has been mostly a descriptive science, devoting insight to small, isolated compartments of a system as a whole – for example, by investigating the influence of individual proteins within the behavior of a whole cell. Thus, the study of the interconnected nature of cellular processes has long been avoided in favor of a reductionist approach. On the one hand this is due to the sheer amount of new challenges that come about when tackling complex systems, whereas on the other hand it has been the common leitmotif in other natural sciences, such as physics, to shed light mostly on well-defined and controllable systems. Those systems are then either small and isolated, or large and homogeneous, so that they can be tackled by applying the laws of statistics. Novel challenges, however, lie in the description of dynamical, mesoscopic, open, spatiotemporally extended, nonlinear systems, operating far away from thermodynamic equilibrium, which are the most important type to understand, as these are the systems that support life.

Although reductionist approaches have been successful in elucidating key processes and key factors of many fundamentally important biological processes, contemporary science is now realizing the importance of wholeness by studying problems of organization. Emergent phenomena arise from the interaction of various units or modules, which are neither resolvable nor understandable through the study of local events or the respective parts in isolation. Hence, traditional reductionist models and methods of cell and molecular biology are not very well suited, and can be incomplete, misleading, or even completely wrong.

Historically, Jan Christiaan Smuts was among the first to formulate a theory of the whole that was hoped to fill the gap between science and philosophy. In his book *Holism and Evolution*, which was published in 1926 [8], Smuts argued that Nature consisted of discrete objects, or ''wholes,'' that are not entirely resolvable into their respective parts. The wholes and parts mutually depend on each other in their functionality, thus forming one organic, unified web of relations, which

comprises matter, life, and mind, and which cannot be accounted for by a reductionistic analysis. Smuts saw his idea confirmed in evolution, regarding Holism as the active driving force towards more perfect wholes or species.

The theoretical foundations of systems engineering were laid some 60 years ago, when the concept of systems theory in biology was proposed during the 1940s by the biologist Ludwig von Bertalanffy [9]. The proposal was further developed during the 1950s by Ross Ashby [10], as a counter-movement against reductionism in science. In the sense of holism, von Bertalanffy emphasized the need for a study of the informational organization within real, open systems. The assembly of such inter-related elements then comprise a unified whole, which in turn can show new emergent properties.

In 1948, the mathematician Norbert Wiener established the field of *cybernetics* [11] as the science of communication and control of systems in regard to their environment. Cybernetics is closely related to systems theory, using the same concepts of information, control, or feedback. However, whereas the former focuses on systems function for providing regular and reproducible behavior, the latter deals more with system structure. Even so, both terms are often used in conjunction, for both structure and function cannot be understood as separate entities.

Today, biology embarks on systems thinking in two different ways. One way is to regard Systems Biology as a new way toward integrating information from different organizational levels, starting from DNA to proteins via signaling pathways to functional modules, into the context of a holistic organizational view [12]. The primary goal of the second view on Systems Biology is to establish a conceptual framework and working methodologies for the augmentation of knowledge on biological phenomena by combining systems theory and molecular biology: ''Systems Biology is not a collection of facts but a way of thinking'' [13]. This view has already been shared during the late 1960s by Mesarović, who predicted that Systems Biology would be an established field of science as soon as ''... biologists start asking the right questions'' [14]. Put differently, biologists need not recast facts already known from molecular biology in a different language, but they need to ask questions based on system-theoretic concepts [15].

Both approaches share the extensive need for high-quality, quantitative biological data obtained through extensive experimental measurements, and this is the reason why the use of systems theory in biology has gained momentum during recent years. New techniques can provide the necessary amount of quantitative data for the establishment of appropriate holistic models of cellular processes. Eventually, those new experimental techniques will lay the foundation for the integration of mathematics, engineering, physics, and computer science into biology, to permit an understanding of the range of complex biological regulatory systems at multiple hierarchical and spatiotemporal levels of cellular organization [16].

2 What Is Systems Understanding?

The word ''system'' derives from the Greek $\sigma\upsilon\sigma\tau\eta\mu\alpha$, and is composed of the prefix syn, which means ''together,'' and the root of histanai, meaning ''cause to stand.'' A *system* is defined as ''... the assembly or set of inter-related elements comprising

a unified whole that is distinct from its environment,'' and can be hierarchically organized and made up of other subsystems or modules, which allows the construction of a complex entity from simpler units. For example, organelles such as mitochondria constitute distinct subsystems within the organization of a cell. The subdivision of natural entities into systems is an abstract construct. Systems *per se* do not really exist in reality; rather, they are defined as a set of elements interacting over time and space.

Systems theory denotes the transdisciplinary investigation of the abstract organization of phenomena, independent of their substance, type, or spatiotemporal scale of existence [17]. The goal of systems theory is to study emerging properties arising from the interconnectedness and complexity of relationships between parts. Such theory argues that however complex or diverse a system is, there are always different types of organizational structures present, which can be represented as a network of information flow. Because these concepts and principles remain the same across different scientific disciplines such as biology, physics, or engineering, systems theory can provide a basis for their unification. The systems view distinguishes itself from the more traditional analytic approach by emphasizing the concepts of system–environment boundaries, signal input–output relationships, signal and information processing, system states, control, and hierarchies. Albeit systems theory is valid for all system types, it usually focuses on complex, adaptive, self-regulating systems which are termed ''*cybernetic*.''

Elegant, simple, and globally valid models are rare in biology as compared to other fields of science. Few examples exist where a function can be attributed to the workings of a single small molecule or few proteins, as in the case of hemoglobin for the transport of gases in the bloodstream, or bacterial chemotaxis [18].

In general, many genes and proteins are involved in cellular responses to external stimuli. In general, biology follows a reductionist approach by investigating small, isolated parts of a cell, tissue, or organism; typically, biology tries to deduce biological phenomena from molecular behavior, which often results in a simplistic ''one gene for one function'' approach. However, since genetic analysis has shown that the genotypes of different species are mostly identical, it would appear that it is the signal processing stage(s) on the way from the genome to the phenotype – in other words, an ever more elaborate regulation of gene expression [19] – which carries the subtle particularities in the respective genetic codes.

As a consequence, biological phenomena should be explained within the vocabulary of system theory, such as amplification, control, adaptation, sensitivity, autoregulation, and error correction, taking an holistic view of the system under consideration [20]. In short, Systems Biology is required to uncover the laws of the whole that cannot be inferred by delving deeper into the details.

The systematic investigation of biological matter comprises the understanding and control of system structure and dynamics in the sense of systems theory and cybernetics, respectively [1].

System structure denotes the identification of the static connection topology and regulatory relationships within the network of genes, proteins, and other small molecules that constitute the signal transduction and metabolic pathways, as well as the physical structure of organisms or cells. Experimental techniques to elucidate

the cells' global system structure include for example DNA microarrays [21], deep sequencing [22], and protein–protein interaction screens via yeast-two-hybrid [23] or split-ubiquitin approaches [24]. System dynamics refers to the qualitative and quantitative evolution of the above network over time. Dynamics include the temporal variation of molecule concentrations, as well as the structure of the network itself. Examples of experimental techniques to study cellular system dynamics include fluorescent imaging techniques to monitor molecular dynamics and interactions on the level of the individual cell [25]. Moreover, mass spectrometry or transcriptomics can be used to investigate the collective behavior of the proteome or gene expression in cells over time.

3 Why Are Biological Systems Different?

While biological systems must still be based on the laws of physics, chemistry, and thermodynamics [26], biology also incorporates the notion of organismal function. This notion represents the need for survival and reproduction, as well as the possibility to evolve in and adapt to changing environments, and it is this inherent purpose that distinguishes biology significantly from all other natural sciences [27]. Moreover, there is no distinct separation of information storage and regulatory units. Genes, for example, can regulate their own expression by gene splicing [28].

Classical physics views the emergence of every effect to be determined by a cause residing in the past. Biological systems, on the contrary, are teleonomic – that is, they are oriented towards a state in the future [17]. It was the insight of cybernetics that purposeful activity can be described through the use of circular mechanisms, where the effect equals the cause. The simplest example of a circular logic is a feedback loop, in which the output of the system is fed into its input again (see Sect. 3.2).

Having these prerequisites in mind, the concepts of systems theory and cybernetics should be utilized, if there is a desire to establish successful formal mathematical models in biology. The introduction of circular causalities has far-reaching consequences on the general design and global properties of biological systems, such as robustness, complexity, and control, all of which are discussed in detail in this chapter.

3.1 Biological Complexity

In order to comprehend the challenges in Systems Biology, it is important first to understand the origin of the complexity encountered in Nature. For once, the complexity within a system does not necessarily come about as a consequence of the number of its component. In physics, the macroscopic state or the dynamics of objects are often well described by a few parameters or simple mathematical equations. For example, although the temperature of an object comes about due to the thermal agitation of its individual atoms, the average energy of the atoms can be ascribed to a single, macroscopically measurable, quantity.

A complex system is one in which the laws that describe its behavior are qualitatively different from those that govern its units, such that new features emerge when moving from one temporal, spatial, or organizational scale to another. The science of complexity is about revealing the principles that govern the ways in which

these new properties appear [29]. Thus, it is rather the organization of the system into irreducible, heterogeneous parts that are highly structured and hierarchically organized on various spatiotemporal scales, which makes a system become complex or, in other words, ''complicated'' [30].

The notion of complexity is not well defined amongst the various science disciplines. Information theory usually characterizes complexity as the amount of information needed to optimally predict the behavior (or state) of the system based on entropy measures [31]. Adami uses sequence complexity of biological genomes to define the amount of information stored about the environment [32]. Another viable approach toward defining the complexity of biological systems may be the identification of the topological structure at a higher level of large-scale organization in terms of hierarchically organized networks [33].

Complex systems can show the emergence of ordered macroscopic behavior, termed *self-organization*. While biological systems are ideal candidates to demonstrate self-organization, the applicability of its principles to elucidate biological phenomena is, to date, rather limited. This is due mainly to the fact that biological systems require a re-definition of complexity that is quite different from that in physics. Biological systems are heterogeneous, modular, highly structured on multiple irreducible spatiotemporal scales, and also self-dissimilar, with each entity usually having several functional and regulatory properties. Nevertheless, the individual components ''. . . interact selectively and nonlinearly to produce coherent rather than complex behavior'' [34].

One explanation for the emergence of self-organization is the so-called ''slaving-principle.'' Under certain conditions the global, macroscopic behavior of a system is governed by a few, slowly evolving state variables, which ''slave'' all other dynamics. In this way, the relevant degrees of freedom are largely reduced, allowing the system to find its own structure – that is, to self-organize [35]. One use of the slaving principle is the analytical description of lasers, explaining the spontaneous, synchronous light emission by the atoms in the lasing medium. Another framework for the explanation of emergent phenomena in multibody systems is the notion of self-organized criticality (SOC) [36], which was acclaimed to explain the frequent occurrence of long-tail distributions in many natural phenomena. SOC denotes the ability of open and dissipative systems to display critical – that is, scale-invariant – behavior even in the absence of external pressures. However, whilst this theory did not live up to its promises of being universally applicable, it is applied in earthquake, forest fire, or avalanche models.

Carlson and Doyle introduced the theory of highly optimized tolerance (HOT) [37], which accounts for the intrinsic design of biological and engineered systems. Such theory reflects the behavior of such high-throughput, high-density systems, which are faced with limited resources. These systems show a high tolerance (i.e., robustness) against environmental parameter fluctuations, with the robustness being achieved at the cost of a high degree of complexity through the addition of control units to the system. Resource constraints then call for an optimized trade-off between fault tolerance towards frequently experienced perturbations and fragility to rare, yet possible, events. Thus, there exists a ''conservation law of robustness'' [38].

The complex design of the living cell is often compared to the make-up of computers or today's commercial aircraft. These

man-made machines have a bewildering complexity, and no man alone understands all the parts and their interplay in complete detail. Many of the strongly interacting and irreducible complex functions are there to automate, control, or back-up operations.

These definitions elucidate the main differences between systems describable by SOC or HOT; the former theory relates to scale-invariant and self-similar features, while the latter deals with self-dissimilar structures and sensitivities (dis-)appearing on each scale of observation, such that the degrees of freedom in these systems are irreducibly many.

3.2 Global Properties of Biological Systems

Typically, systems thinking seeks universal properties for biological systems, linking the emergence of complexity to general design features. Such features can be used as a guideline for abstract mathematical modeling by examining the principles that are commonly shared between diverse species. These properties, as discussed below, show how biology and engineering converge on a systems level view.

3.2.1 Robustness of Biological Systems

The fault tolerance or robustness of biological systems denotes the maintenance of specific system functionalities, even in the presence of fluctuations or a change in environmental parameters [39, 40]. In biology, this refers to the concept of homeostasis and the stability of developmental control. Robustness is achieved through the incorporation of regulatory control loops into a system, thus shielding or buffering the desired system functionality from environmental influences. Robustness is a relative system property: in order to maintain a certain equilibrium, other properties must evolve and adapt. There is, therefore, a need to define which cellular functions change, and which resist as a reaction towards a disturbance.

Most of the ''complicatedness'' encountered in biological systems is a direct consequence of the implementation of control schemes, rather than the core function itself. In particular, if robustness is to be achieved for a wide range of disturbances, the control must have an equally increasing variety, as stated in the ''Law of Requisite Variety'' [10]. Consequently, complexity serves simplicity in the sense that the complicated control schemes are hidden underneath the simple, yet reliable, output [41].

Robustness in biological systems is usually achieved by various strategies, including redundant or functionally overlapping regulatory pathways [42], feedback loops to regulate signal responses [43], or checkpoints within the cell [44, 45]. All of these control elements add to the total number of components regulating cellular signaling elements, while keeping the number of phenotypic expression levels low [20]. Cellular checkpoints play important roles, such as in mitosis [45, 46], in the cell cycle [47], and during the early stages of embryo development [48]. Bacterial chemotaxis is an intensively studied example of a robust behavior due to multiple feedback control [49, 50]. Examples of genetic buffering to achieve robustness are shown by the fact that only about 20% of genes in budding yeast are essential for viability [51]. To a certain extent, genetic buffering can safeguard against environmental changes or genetic defects. The simplest bacteria (e.g., *Mycoplasma pneumonia*) may have only a couple of hundred genes yet, despite their unexpectedly complex transcriptome organization (which includes antisense

transcripts, alternative transcripts, and multiple regulators per gene [52]) they survive only within a narrow band of environmental parameters. In contrast, the bacterium *Escherichia coli* possesses about 3000 genes, and is able to survive under a variety of environmental conditions [30] by activating additional control schemes under environmental stress.

A further consequence of robustness is that dynamical behavior in biological systems is coupled less to the parameters themselves, and more to the overall system structure. Previously, Von Dassow [53] showed that robustness was the simplifying criterion for determining the correct topology of the segment polarity network, producing a highly robust patterning over a large range of parameter variation.

Robustness in complex systems comes at the price of some fragility due to tiny, yet rare, events. Indeed, a small rearrangement of some cellular signal pathways can often lead to a spectacular failure – that is, impaired cell death. For example, although cancer may be caused through an accumulation of mutations in the genetic code over the human life span [54], the occurrence of cancer during the reproductive years of a human is rather rare. It appears that Nature can cope well with this trade-off for the benefit of robust and reliable ''normal'' functioning.

3.2.2 System Adaptation and Control

Adaptation is the ability of a system to accommodate for varying external input stimuli or disturbances in order to gain and maintain the correct and optimized output; this is usually achieved with the help of feedback control. A well-studied example of adaptation using integral feedback control is chemotaxis, in which bacteria adapt their movement not to the absolute level of pheromones but rather to the chemical gradient, only [18, 50, 55]. Another example is the activation of heat shock proteins in *E. coli* under temperature stress [56], which function as chaperones for correct protein folding and the prevention of unwanted protein aggregation.

Control is a recurrent design encountered in natural systems, which effectively increases robustness, with the goal of keeping certain values within predefined physiological limits. It is possible to distinguish between feed-forward and feed-back system controls; the former is an open-loop sequence of predefined actions triggered by a certain stimulus, working dependably only within strict ranges of input stimuli, yet simple in design [57], while the latter employs a closed-loop design, which feeds part of the output signal back into the system input. Depending on the sign of the feed-back, positive feed-back (or autocatalysis) will amplify the output signal, enhancing reliable sensitivity toward cellular decision derived from noisy input signals [58], whereas negative feedback usually stabilizes the output around some desired value by opposing any changes caused by disturbances. For example, a combination of feed-back control schemes can be used to stabilize the receptor presentation of a cell to increase its sensitivity toward a broad range of external ligand concentrations; this was demonstrated in the case of erythropoietin receptor signaling [59].

Although both types of control are successfully employed in intracellular signaling [60], both also bear fragilities. Autocatalysis can lead to self-sustained, unrestricted signal amplification, as observed in uncontrolled tumor growth [61], while ultra-stable homeostasis regulation can cause large, possibly disturbing,

transient signals in response to external fluctuations [38].

Most often, biological systems use combinations of open- and closed-loop designs, basing their control action on the absolute, accumulative or differential value of the input stimulus, in this way balancing both sensitivity and stability.

3.2.3 Modules and Protocols

Modules are subsystems characterized as mostly self-contained entities with fixed interfaces for external communication. They are evolutionarily detached, possessing their own identity, and have many internal – but only few external – links for information and matter exchange [27].

Modularity seems to be an important concept in biology, probably stemming from the evolutionary pressure for optimal flexibility and the low chance of damage spread. Spatially distinct entities, such as organelles in the cell or metabolic units, appear to be a recurrent scheme found throughout various organisms. The gene regulatory network in cells seems to be organized in modules [62]. Interestingly, modular design is a commonly found principle in modern engineering, as it enables the independent development and testing of units before their integration into a common system. The benefits comprise savings in developmental and maintenance costs, as well as the prospect of graceful degradation rather than catastrophic failure, as errors are usually restricted to a single module.

If the modular design of biological systems is a governing principle, this opens up new possibilities for the simplification of *in vivo* and *in silico* experiments. Modules would provide fixed levels of detail and size, which are easily abstracted once their functions are known. Moreover, with the help of these core modules it would be possible to build ever-more complex models, without the need for segmentation of every level of detail [62].

Modules communicate by using *protocols*; these are fixed, commonly agreed-on rules that standardize communication with their respective surroundings. In addition, protocols ensure error correction, cellular coordination, and also evolvability through the possibility of adding new functions to a particular module. Protocols have been shown to be an efficient means for ensuring the hierarchical organization of complex systems through their integration of different layers, thus reducing the costs of information transmission [63]. For example, negative feed-back is a powerful module for the establishment of homeostasis. Similarly, gene regulation, membrane potentials, or signal transduction pathways can all be regarded as protocols which are utilized by different biological modules, such as the DNA, ion channels or kinases, and phosphatases [38]. Elaborate feed-back control protocols have been robustly established for the spatiotemporal development of various species [60, 64]. An extensive genome-wide analysis of prokaryotic cell-cycle progression has revealed a hierarchical control structure with three to four master regulatory proteins acting in a coordinated fashion [65].

Interestingly, the use of protocols on a cellular level has striking similarities with the way that modern computers communicate via the internet. The internet employs many protocols for persistent communication that can be ordered into a general hierarchy, starting from the data link via the network and transport to the application layer. Popular examples of each hierarchy are the reverse address resolution protocol (RARP) for the IP (internet protocol) look-up and search of the communication partner, the IP for the routing

for the appropriate subnet of the data, the TCP (transmission control protocol) for the secure, error-free data exchange and, finally, the application layer such as the HTTP (hypertext transfer protocol) or FTP (file transfer protocol) for retrieving a web page or downloading files. The fact that these protocols have been used for decades, irrespective of the rapid and ongoing evolution of computer hardware and software, highlights their importance.

4 Systems Biology Modeling

The word "model" derives from the Latin *modus* – that is, "manner" or "measure," and refers to the concept of representing causal relationships from real systems in the language of mathematics. This mapping often involves simplifications of the original systems, with the hope of gaining predictive power on experimental results and explaining functional design principles. Life is an emergent property stemming from the interaction among molecules, and cannot be reduced to the individual properties of the molecules. Modeling tries to infer and predict the relations between molecules in terms of causation – that is, it tries to establish explanatory relationships of the spatiotemporal changes of matter [66].

Modeling and simulation have become indispensable tools for gaining insights into natural systems. In biology, they help to bridge the gap between theory and experiment. Often, biologists are faced with the dilemma that experiments do not provide sufficient data for theoretical interpretation, while at the same time, clues for new experiments are missing as a consequence of lacking hypotheses [67].

Experimental results require correct mathematical interpretation, and model hypotheses require experimental proofs [68]. The process of knowledge generation is iterative in nature, and consists of two feedback loops (see Fig. 1). Here, the experimental part employs high-throughput techniques to obtain quantitative data

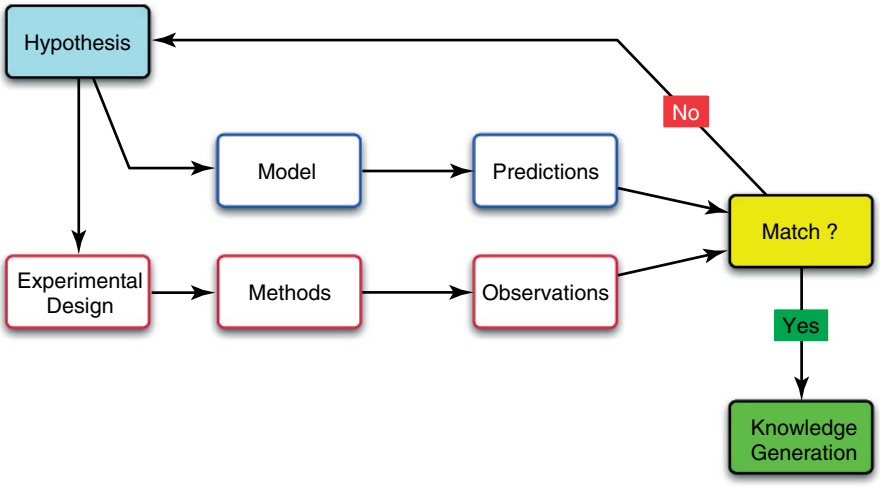

Fig. 1 Schematic to the experiment–modeling loop for knowledge generation in Systems Biology.

relating to the system dynamics; the data acquired are then compared with predictions obtained from the mathematical model. In the case of failure – that is, in the case of diverging results from model and experiment – this process must be repeated by adjusting the old hypothesis and thus reconciling the model with the new results from the experiment.

As an example of this cycle of model building, experiment, and model refinement, Ideker *et al*. [69] have recently developed an integrated approach to construct, test, confirm and refine the simulation of the yeast galactose utilization network, through the combined numerical simulation and analysis of networks with systematic experimental perturbation and global measurements.

The major guideline for optimal modeling should be the concept of Occam's razor, which states that models should be void of any redundant information. Yet today, this paradigm needs also to be viewed from the opposite aspect – models must not be oversimplified, so as to miss the essential clues of functionality of real life systems. In particular, it is important to note that modeling in biology must establish different concepts from those in physics. The paradigm of nonlinear science is that even simple systems can lead to a dynamically rich behavior. Despite the beauty of the simplicity of this idea, however, it must be realized that living systems regulate their dynamics differently – that is, through a complex make-up of interconnected regulatory functions, a fact which is often neglected or ignored [70]. Models in biology are usually heuristic; they arise embedded in the process of biological experiments, and are coupled tightly to them. Information is, for example, deduced from a perturbation analysis of the experimental system, and thus contains assumptions regarding the causality and the passage of time [70].

In general, experimental data must be comprehensive with respect to four aspects [57]:

- **Factor**: the need to capture the behavior of all important target factors, such as genes and proteins that play decisive roles in the experimental system under consideration.
- **Item**: this refers to the simultaneous measurement of the necessary sets of variables that are required for reliable hypothesis building, such as transcription level, molecule concentration, or spatial location.
- **Time and space**: this refers to the need for a sufficiently high sampling rate of the experimental data to obtain a reasonable resolution of the spatio-temporal dynamics.
- **Repetition**: experiments need to be repeated to obtain a statistically reliable estimate for the biological variability and other sources of error induced by the experimental set-up.

The construction of a valid working model of a biological system from experimental data can be approached from two directions:

- **Bottom-up modeling**: This is based on the integration of established biological knowledge on the dynamics of the relevant biological components of the system or regulatory network under consideration. This attempt is useful when most of the reaction partners are known and their interaction dynamics are understood. The research goal is the establishment of an accurate computer simulation that allows for: (1) the analysis of the system dynamics;

(2) the scan of parameter ranges that are unattainable in experiments; and (3) the prediction of unknown functionality or interactions [57]. Attempts at bottom-up modeling include the λ-phage decision circuit [71] or the data-driven simulation of a cancer cell [72].

- **Top-down modeling**: This approach attempts to apply statistical analyses to data from high-throughput experiments derived, for example, from DNA microarrays. Data mining techniques search for clusters of coexpressed genes as a consequence of cell state or an external perturbation, such as the knockout or overexpression of certain genes.

The working assumption for these methods is that coexpressed genes also share common relationships with respect to other biological processes. For example, Müller *et al*. [73] showed, from clustering analysis, how pluripotent stem cells are under tight control by specific molecular networks across species.

The result of a cluster analysis is the construction of an interaction network of genes or proteins, the topology of which can hint at biological reasonable organization principles [33, 74, 75]. While top-down modeling is well in line with the need for holistic approaches, it has been criticized for violating individuality and locality in the cell. Although cellular stimulus–response patterns are highly coordinated, these patterns emerge from individual protein–protein interactions, which cannot be deduced from the high-throughput data. Hence, the question of gaining new biological knowledge on individual genes or proteins from this approach often remains open.

Although knowledge-based bottom-up approaches of modeling possess a certain appeal to biologists, successful modeling needs to overcome the gaps in understanding. Most protein–protein interactions remain unknown, and the underlying physics of interacting molecules require better attention. Until now, the influence of high molecule concentrations in the subcompartments of a cell environment on reaction rates is largely unknown, but the spatial distribution of reactants and their molecular crowding may well affect the reactions and their rates [76, 77]; an example of this is when explaining the symmetry breaking process of mitosis [78]. One particular challenge arises when trying to identify the relevant components, as this is very difficult due to the vast number of combinations of active molecules. Hence, it remains unclear as to how this approach can be scaled up to large networks on the cellular or even tissue level. Indeed, this situation presently poses major difficulties for detailed mathematical modeling in biology.

Regardless of the approach taken, there is a need to define the level of model abstraction, complexity, and spatiotemporal scale of the system under investigation. Cellular processes have characteristic time scales that range from milliseconds for individual protein–protein interactions to minutes for phosphorylation events, up to hours, days, and years for changes in gene transcription, cell growth, and gene mutations, respectively. This has important consequences for the appropriate choice of model detail level. Processes that occur much faster than the time scale of observation can be assumed to be instantaneous, while slow processes can be assumed to be quasi-static. As a consequence, the level of abstraction leads to certain dynamics being modeled and simulated in detail, whereas for other parts the details can be neglected – for example, due to low sensitivity towards specific

parameter values. Recently, Busch *et al.* [79] have used time-scale separation between fast protein signaling and slow transcription dynamics to infer a dynamic decision network for hepatocyte growth factor-induced migration of keratinocytes. On the time-scale of observation, all protein concentrations were assumed to be in quasi-equilibrium, such that it sufficed to focus on the change in gene regulation as the decisive element of regulation.

The major goal of theory and model simulations is the reverse engineering of biological systems, which is "... the process of analyzing a system to identify its components and their interrelationships and create representations of the system in another form or at a higher level of abstraction" [80]. Despite major efforts in this field [81], biological systems still pose a major challenge towards reverse engineering, due to the hidden complexities, inherent robustness, and possibly suboptimal design of functional units. Circular causality makes it difficult to distinguish between cause and effect from biological data, and all of this imposes ambiguities when attempting to deduce the correct biological "wiring diagram" from the experimental data. It is hoped that systems thinking in biology, with its concepts of robustness, hierarchy and modularity, in addition to the necessary protocols, will provide detailed bottom-up models with testable hypotheses for model discrimination [39].

4.1 Network Biology

Network biology is the study of the static organization of life as networks of biological entities such as DNA, proteins, RNA, and metabolites [33, 82].

The relationship between these entities is depicted graphically, the result being a set of nodes connected by edges. Usually, the nodes represent the state variables of the system (such as the molecule concentrations), while the connections define the interaction between the nodes. The networks are characterized by their connectivity, path length, and clustering distributions. Connectivity indicates how many neighbors each node has on average, while the path length denotes the average separation between arbitrarily chosen nodes, and the clustering coefficient is a measure of the grouping tendency of the nodes.

Network biology attempts to discover universal design and organization principles, which govern the functioning and evolution of intercellular and intracellular networks [83, 84]. Biological networks appear to share certain topologies that are best described as scale-free networks; these possess few highly connected nodes, termed *hubs*, while many nodes share only a few connections. Hence, the connectivity distribution follows a power-law: $P(k) \propto k^n$, where k and $P(k)$ denote the connectivity and distribution, respectively, and n is the connectivity exponent. These networks appear if the new nodes have a preferential attachment to already highly connected hubs [85, 86], although, interestingly, this type of complex network also emerges in other aspects of life and society, such as the worldwide web and social interaction webs [87].

Network architecture and task are closely linked. By comparing the topology of the regulatory network in *E. coli* with the call graph of the kernel of the Linux computer operating system, Yan *et al.* showed that both networks show a hierarchical layout, despite having fundamentally different design principles [88]. The *E. coli* network is optimized toward robustness, with few global regulators at the top and many downstream targets, whereas the

Linux kernel is designed for code efficiency and the re-use of software modules at the cost of robustness, and has many regulators controlling a small set of highly connected generic functions. By using high-throughput methods, it is possible to draw ever more-detailed interaction webs of protein–protein, metabolite, and gene transcription networks. Experimental results obtained with yeast [89] and *E. coli* [90] seem to support the view that both metabolic and genomic regulatory networks show a hierarchical organization with few recurrent subnetworks, termed *motifs*, which hint at the existence of elementary regulatory units [5, 74]. While modules are discrete functional units that are semi-detached from the whole system, motifs comprise a set of genes or metabolites that form recurring, significant patterns of interconnections, that are inseparable from the remainder of the system [87]. Subsequently, Ravasz *et al.* showed the metabolic networks of 43 distinct organisms to have a modular organization that was interconnected in a hierarchical manner – a system-level cellular organization that might be generic in nature [91].

The need to investigate the interconnections of genes–proteins and proteins–proteins stems from the fact that the genome of various species is quite similar, despite it having been argued that the evolution of ever more-complex species is correlated with an increased functional connectivity between a constant number of genes. Motifs might provide a means to increase the interconnectivity of existing proteins, in this way creating new functionalities. As noted by Ravasz *et al*:

> *"This is likely one of several reasons that the apparent complexity of organisms can increase so markedly without a corresponding increase in gene number. An attribute of proteins encoded by the human genome is that they have a richer assembly of domains than do their counterparts in invertebrates or yeast, and indeed the assortment of domains into novel combinations is likely an important aspect of genome divergence" [92].*

Today, network analysis has opened up new avenues in the global analysis of diseases, and their mutual connections. Such analysis allows for the identification of the genetic basis of disease [93], and can also reveal novel gene association overlapping across common human disorders, which in turn helps to unravel the general patterns of human diseases that are not clear from studies of the individual conditions [94].

4.2 Dynamic Network Models

Despite the intuitive appeal of network biology for the structure identification of biological systems, it is limited in the sense that it does not include the temporal dynamics of the system. Life is an emergent property of the interaction of cellular proteins. Hence, the dynamic interplay occurring via chemical reactions and changing cellular protein numbers is essential to the cell function, and constitutes the essence of many modeling approaches in Systems Biology. Boolean networks are an abstract, yet relevant, approach towards the inclusion of a temporal evolution of large-scale networks, as they offer a qualitative modeling approaches to build and analyze simplified, but still rigorous, dynamical models. Boolean networks are used for the elucidation of

large-scale dynamic protein signaling networks, where it is assumed that the detailed inclusion of, perhaps, concentration gradients or the stochastic effects of protein concentrations, can be neglected in favor of including a large number of players [95]. Each node of a Boolean network can assume an ON/OFF value. Then, the state of each node at the next time step $(t+1)$ is deduced deterministically from a logical Boolean function (i.e., AND, OR, NAND . . .), based on its current state and external input [96]. Boolean networks have been used successfully for the analysis of protein and gene regulatory networks [97], for the modeling of the epidermal growth factor receptor (EGFR)/ErbB signaling pathways [98], and apoptosis [99].

4.3 Reaction–Diffusion Models

Biochemical reactions can be considered as the most fundamental processes in cells, wherein the concentrations of the reacting species change subject to the other molecule species involved within the respective reactions. Ordinary differential equations/partial differential equations (ODEs/PDEs) are a natural choice for the mathematical description of system dynamics with continuous system states in time and space, respectively [6, 100]. They are, therefore, ideally suited to describe changes in concentrations in biochemical reactions, being the most widespread formalism to model systems throughout the various scientific domains. In biology, they are widely used to describe the time and space course of molecular concentrations. In this case, differential equations relate the rate of change of a variable (protein substrate) to the current state of other variables (reactant), wherein the interaction between the various reactants is modeled through functional and differential relations. There is a vast literature available on modeling biochemical reactions based on differential equations, especially in the context of metabolic processes (cf. Ref. [100] and references therein).

In general, it is possible to distinguish between purely time-dependent ODEs, and more general PDEs, which additionally include spatial dimensions. A PDE describing the spatiotemporal evolution of a system has the general form:

$$\frac{\partial \psi_i(\mathbf{r},t)}{\partial t} = \underbrace{f\left[\psi_i(\mathbf{r},t)\right]}_{\text{temporal evolution}} + \underbrace{\nabla \mathbf{D}_i(\mathbf{r},t)\nabla\psi_i(\mathbf{r},t)}_{\text{spatial Diffusion}}, \tag{1}$$

where $\psi_i(\mathbf{r},t)$ denotes the respective system state variables of the various molecules labeled by the subscript i. The above equation reduces to an ODE in the absence of spatial diffusion, that is, $\mathbf{D}_i(\mathbf{r},t)=0$. The term $f[\psi_i(\mathbf{r},t)]$ denotes the respective synthesis rates, depending on the various concentrations of $\psi_i(\mathbf{r},t)$ and possibly external signals. It is often comprised of the various, mostly nonlinear, interaction functions between the system reactants based on the ''law of mass action,'' which leads to, for example, Michaelis–Menten-like enzymatic degradation or the Hill-type cooperative activation [6, 100]. The organization of interacting molecules within a cell implies the concept of pathways, in which information processing in the cell is organized. In terms of mathematics, such a biochemical network is then represented as a system of coupled differential equations, as shown above. It is those nonlinear interactions that are essential for biological systems to show nontrivial behavior such as multistability, hysteresis, or oscillations [43].

A stability analysis of the system of coupled equations is usually carried out to unravel the qualitative behavior of steady-state solutions and their stability, as well as the occurrence of periodic solutions in space and time [101]. As a matter of fact, the dynamic behavior of ODE systems depends heavily on the reaction parameters of the underlying chemical reactions. Therefore, the estimation and identification of parameters from experimental data [102, 103], as well as the optimal experiment design to yield a maximal amount of new biological knowledge [104, 105], is a field of active research in Systems Biology. Due to the nonlinear interaction terms $f[\psi_i(\mathbf{r}, t)]$, an analytical solution of the differential equations is usually impossible, and solutions must be found via numerical integration. A variety of numerical integration algorithms for ODEs and PDEs can be found in the literature [106]. In addition, various computer software packages have been developed specifically for the simulation and bifurcation analysis of nonlinear systems, such as GEPASI [107], DBSolve [108], Cell Designer [109], or a Matlab extension such as SB toolbox [110] or Potters wheel [111].

The great success of modeling is its predictive power. From model simulations, it is possible to obtain a better insight into the regulatory logic, and it is also possible to perform experiments in the computer that otherwise were impossible. A good example of this is the mitogen-activated protein cascade model, as originally proposed by Huang and Ferrell [112]. Here, the pathway is highly conserved and implicated in various biological processes, conducting signals from the membrane to the nucleus. It is composed of three kinases that sequentially phosphorylate and activate each other. By converting and simulating the rate equations into a system of coupled ODEs, Huang and Ferrell showed that this particular pathway architecture could be used to convert a graded response into a switch-like robust output, appropriate for mediating processes such as mitogenesis or cell fate induction.

The spatial aspect of the equation enters the description due to the inclusion of a diffusion term, where the diffusion term $\mathbf{D}_i(\mathbf{r}, t)$ depends on space and time in general. Alan Turing was the first to point out such reaction–diffusion systems as a possible explanation for morphogenesis in natural systems, when he showed – theoretically – that a system of reacting and diffusing chemicals can evolve spontaneously to a spatially heterogeneous state as a response to an infinitesimal small forcing [113]. These pioneering studies led to a new branch of research, and many so-called ''Turing systems'' have been proposed (though not finally proven) to account for pattern formation in developmental biology. Examples include an explanation of the pattern formation on snail houses [114], the modeling of *Drosophila* embryogenesis [115], or the coating patterns of animals [100]. The importance of spatiotemporal inhomogeneities in molecular concentrations and molecular crowding in signal transduction pathways has also recently been acknowledged [77, 116]. As a consequence, cellular functions may also employ active transport mechanisms to sustain reliable cellular processes [117, 118].

The above ODE formalism fails in the case of low molecule concentrations [119], as the discrete and random nature of the individual, elementary reactions between molecules then becomes non-negligible, showing an impact on a macroscopic scale as fluctuations in the molecule concentration over time [120]. This chemical noise becomes most significant in the regulation

of gene expression, which is usually accompanied with low copy numbers of mRNA transcripts and the genes themselves [121]. Stochastic effects in cellular pathways have been attributed to cause phenotypic diversity in isogenetic populations of cells [122], or to play a major role in the lysis-lysogeny decision circle of the λ-phage [123].

4.4 Holism versus Reductionism: The Global Dynamics of Networks

Currently, most Systems Biology approaches follow either a top-down or a bottom-up approach. However, while both have their unique motivation, they each also hold certain criticisms. Typically, a top-down approach allows for the holistic, unbiased view of cellular events, albeit at the expense of limiting detail level, thus violating the individuality and locality that are important in the understanding of cellular processes. The bottom-up approach aims at a detailed mechanistic and causal understanding of biochemical networks, but it is not yet been determined how such studies of ''isolate'' signaling pathways can be scaled up to the cellular or even tissue and whole organism levels. As an interim solution, Huang and Ingber [20, 124] have proposed the study of cellular behavior on the level of global network dynamics to reveal any higher-order, collective behavior of the interacting genes and proteins. These authors have described global network organization in terms of a state space, spanned by the expression levels of the whole genome, as well as attractors providing a mathematical and molecular basis for an epigenetic landscape. Genome-wide expression levels cannot take up arbitrary values; instead, they are tightly coupled to respective cell states – that is, to the various cell types and distinct, stable phenotypic states in a multicellular organism. Such stable network configurations are referred to as *attractor states*. Huang and Ingber have argued that the attractors naturally capture the essential properties of cell behavior, such as the mutual exclusivity of cell fates, robustness, and all-or-none transitions in response to a large variety of signals. As a proof of principle, it was shown that HL60 cells are capable of following different routes into the same differentiated state. Subsequently, upon the application of two different stimuli, separate transcriptome response transients were created which, nevertheless, settled onto the same gene expression pattern after several days [125].

This formal framework on the orchestrated role of cellular gene network dynamics could serve as a potential explanation for stem cell differentiation and the reprogramming of differentiated back into pluripotent stem cells [126], or even cancer progression. The accumulation of genetic mutations over time would distort the attractor landscape such that it would eventually lead to an altered cellular response within the same cellular context [124]. While the above suggestions are intuitive and appealing, their value must still be proven with respect to knowledge gain on the level of individual genes and proteins as potential targets for cell control and intervention.

4.5 Modeling Resources and Standards

The development of a standardized computer infrastructure is of utmost importance to manage the increasing amount of knowledge relating to biological systems. This supports the effective utilization of resources and the exchange of models,

ideas, and data. An integrative software tool for Systems Biology should comprise the following features for proper systems understanding [67]. From a mathematical point of view, these tools need to support different simulation algorithms, such as deterministic and stochastic ODE and PDE solvers, and should include analysis algorithms for parameter estimation and model discrimination.

From an experimental point of view, a package should support a standardized modeling language, while in terms of software the simulation package should run independently from the computer platform, preferably on a computer grid or cluster environment.

A number of simulation packages are currently under active development, most of them freely available. As a consequence of the complexity encountered in biological systems, there is no integrative software package presently capable of handling all phenomena in signal transduction, metabolic pathways, or spatiotemporal simulations. Some of the currently most versatile packages among many are ECell [127], Virtual Cell [128], Cellware [129], Cell Designer [109], or Smart Cell as simulation packages for spatiotemporal simulation [130].

Model building is a complex enterprise, and is usually accomplished in collaboration with scientists from different research institutions. Serving the need for seamless information exchange of computer-based models, progress has been made to define standardized biological ''wiring diagrams'' [131] together with a common description language, SBGM [132]. Moreover, common standard mark-up languages based on the XML format help in the unique definition of biological entities, pathways, and events for the rapid exchange of models between experimentalists and theoreticians. The most common of these are CellML [133], BioPAX [134], and the Systems Biology Markup Language, SBML [135]. The Systems Biology workbench extends the above approach by providing a standardized application interface for researchers to exchange not only model data and results, but also the simulation tools themselves [136]. Another challenge lies in the inherent modeling of the stochasticity of inner cell processes.

The SSA (stochastic simulation algorithm) from Gillespie is computationally infeasible with an increasing number of molecules, and the computational power for exact stochastic simulations will be immense. Consequently, different algorithms have been developed, such as to reduce the number of random variables for simulation [137] or to allow larger time steps to be taken as a justifiable error in the respective reaction probabilities [138]. Several program packages for the simulation of stochastic molecular dynamics exist, including StochSim [139] or Stocks [140].

Regardless of the modeling approach of biological systems employed, there is always a need for parameter estimates of at least a few constituents of the model. Because of the sheer amount of parameters required for successful modeling, and the huge number of experiments already performed, the need to organize experimental data has brought about several publicly available databases of molecular properties, interactions, and pathways [141–143]. These provide an invaluable infrastructure for future modeling efforts, enabling the modeler to begin simulations from a certain degree of abstraction [144].

Spatiotemporal modeling will additionally require information concerning the physical structure of its model constituents, as obtained using microscopy. For these needs, the development of an

Open Microscope Environment, which is currently being built as a joint effort among various research institutions, will provide a unified data format and database environment for consistently annotating, storing, and retrieving five-dimensional microscope images (four dimensions in space and time, with additional color information) and exchanging them between research institutions [145, 146].

5 Future Prospects of Systems Biology

The systematic generation and analysis of quantitative experimental data slowly, but definitely, is turning biology into branch of science that is close to engineering. Once the "language of the genes" – that is, their syntax and their semantics – is decoded [147], then theoretical knowledge and experimental expertise will suffice to draft and create synthetic model cells or even organisms from scratch, on which new drugs and cures can be tested *in silico* before their possible assembly *in vivo* [148].

There is indeed the prospect that Systems Biology will change medical practice by allowing the prediction of new combinatorial and/or personalized drugs for diseases that currently are regarded as severe, including Alzheimer's, diabetes, human immune deficiency virus (HIV), or cancer [149]. Indeed, Systems Biology is set to change today's medicine from being responsive to being predictive, preventive, personalized, and participatory; this situation is often referred to as *"P4 medicine"* [150]. For example, viewing cancer from a systemic level as a robust system might provide physicians with a framework for future anticancer strategies [61]. As a consequence, anticancer therapies might be developed from mathematical modeling to first identify the peculiarities of an heterogeneous tumor cell population, and then try to predict and control cell activity by detecting and using specific fragile points for each tumor cell type [151]. As a first success in this line, novel, computationally predicted anticancer drugs targeting the ErbB family of receptor tyrosine kinases have been developed from *in silico* models of signaling pathways, and are currently undergoing clinical trials [152, 153].

5.1 Synthetic Biology

The aim of Synthetic Biology is to (re-)design new or already existing biological parts, devices, and systems with the help of mathematical modeling and engineering approaches; in other words, Synthetic Biology is the technological counterpart of Systems Biology. Yet, progress in Synthetic Biology goes hand in hand with current progress in cellular and molecular biology, as well as genetics and associated fields of engineering and computer sciences. It is hoped that the combination of this knowledge will enable the creation of essentially artificial systems, by employing biological design principles with new combinations of building modules from existing (sub)cellular systems [154, 155]. In fact, the first steps in the creation of artificial genomes and their implantation into host organisms has already achieved a degree of success [156].

Synthetic Biology departs from a component-based approach by viewing a living system as a programmable entity that is composed of interacting modules, each having particular functions and which exchange their information via protocols. The results of current research have suggested that these modules are (possibly) limited in number, albeit their

specific tasks are diverse. Synthetic Biology implies a scientific agenda on a higher level of abstraction towards identifying and categorizing the various module types within the cell and across organisms, and investigating their interactions on the basis of the modular rather than the molecular interaction. From an engineering point of view, this means that modules can be taken out of their current evolutionary context and assembled differently, in the sense of versatile building blocks.

Ideally, Synthetic Biology starts with mathematically inspired designs, such as a system of coupled differential equations with desired dynamic properties, which are then translated into the biological and chemical realities as promoters, enzymes, or metabolites within a cell. Along this line of thought, the workings of switches and bistable systems of stabilizing control loops have been systematically recast in terms of chemical reaction schemes [148, 157]. Different authors have described the fundamental principles of building logic circuits into the language of gene regulatory networks. For example, Tyson *et al.* presented a systematic overview of designs for biological control systems such as switches, sniffers or buzzers, and combined their mathematical description with experimental findings [43]. Likewise, Hasty *et al.* have reviewed the possibilities of constructing gene circuits which serve various functions such as autoregulation, repression, or logical gates [157]. In a first *de novo* design study of a gene regulatory network, Guet *et al.* systematically analyzed the phenotypic behavior of different parameter and topology combinations in a genetic regulatory network [158]. These authors constructed various logic gates, such as NAND, NOR, and NOT IF, through the combination of three nodes together with five promoters. As an important result, it was found that not only the parameter values but also the network topology was important in order to determine unambiguously the computational function of the systems.

Recently, efforts have been undertaken to establish a Biological Information System (BIS) which extends genomic databases with quantitative mechanistic knowledge [70]. The BioBricks foundation (http://bbf.openwetware.org/), founded by engineers and scientists from the Massachusetts Institute of Technology (MIT), Harvard and the University of California, San Francisco (UCSF), have set up a free, publicly available repository of standardized biological parts for the common re-use and combination to design *de novo* functions in living organisms. Bio-Bricks constitute promoters, proteins, RNA-coding sequences, or transcriptional terminators; physically, they are DNA sequences stored on a circular plasmid distributed by the Registry of Biological Parts (http://www.partsregistry.org). By using such data in an integrated form, it will become possible to build refined synthetic systems by checking on key molecules and replacing complex pathways with effective reaction parameters, thus defining biologically meaningful reaction subsets from the large amount of possible reactions.

5.2 Conclusions: Where Are We?

Although, today, Systems Biology is still at the start of a new branch of science, the road ahead is clearly marked and commonly agreed on by many people. Yet, the problems associated with all of the processes lie in the details.

To date, most experimental biosciences are method-driven, and pay much attention to detail and the production,

in extreme cases, of large amounts of data which are out of context. Theoretical sciences, on the other hand, are principle-driven; they neglect important details and thus come up with theorems and models that are "out of this world." As a consequence, research groups must learn to focus on the important areas of their respective systems under investigation, while also determining which part to measure in detail and which part to neglect:

- *Biologists* need to adopt abstract thinking, and to trust in the language of mathematics and necessary simplification.
- *Physicists* must learn to abolish oversimplified thinking [159] and become used to the analysis of strongly interacting systems with many degrees of freedom.
- *Engineers* must gain a deeper understanding of their systems under investigation beyond numerical solutions, for example, in terms of process optimization.

The necessary unification of these various branches of science will, for the sake of the advancement of Systems Biology, make communication skills among scientists ever more important. Previous years of "isolated" research have diversified the scientific vocabulary and methods, making it sometimes very difficult to mediate and discuss interdisciplinary goals and methods. Moreover, all scientists must learn to place their own research into a much wider frame of interdisciplinary research and research teams.

Despite this increased complexity of the systems under study, one of the greatest challenges in Systems Biology is to bridge the gap between detailed kinetic protein signaling models (of up to 100 different molecules) and large-scale -omics approaches, thereby providing – simultaneously – data on thousands of genes or proteins. At present, the means by which the detailed models can be scaled up to include more variables and to bridge several time scales (from minutes to hours or even days) is largely unknown. In contrast, it is very difficult to move from the statistical analysis of -omics data to the prediction of individual protein interactions. Yet, taken together, despite not being able to bridge this modeling gap, there is a clear need for an intuitive understanding of the relevant elements of the system under study. Vilar *et al.* [160] have argued, using the example of the lac operon, that "... even in the 'postgenomic era,' (modeling) will still rely more on good intuition and skills of quantitative biologists than on the sheer power of computers."

Today, it is strongly believed that Systems Biology will promote a unification of the sciences, especially if the scientific community continues to work "holistically" in a joint effort of biology, mathematics, physics, and engineering. Only then might a new generation of scientists with an interdisciplinary training emerge, whose everyday business will comprise not only working in the laboratory but also performing data analysis and model simulations, in addition to exchanging their data and results freely through publicly available databases. With the establishment of new experimental techniques and mathematical tools, and asking the right questions, it is clear that the systems approach to biology will become a major success. Indeed, it will change not only modern life sciences, but also views on life itself.

References

1 Kitano, H. (2002) Systems biology: a brief overview. *Science*, **295**, 1662–1664.
2 Special Issue (2002) Systems biology. *Science*, **295**.
3 Special Issue (2002) Systems biology. *Nature Insight*, **420**.
4 Szallasi, Z., Stelling, J., Periwal, V. (2006) *System Modeling in Cell Biology*, The MIT Press.
5 Alon, U. (2007) *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman & Hall/CRC.
6 Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., Lehrach, H., Herwig, R. (2009) *Systems Biology*, Wiley-Blackwell.
7 Fu, P., Panke, S. (Eds) (2009) *Systems Biology and Synthetic Biology*, AIChE-Wiley Press.
8 Smuts, J.C. (1926) *Holism and Evolution*, Macmillan & Co. Ltd, London.
9 Bertalanffy, L. (1973) *General Systems Theory*, Penguin, Harmondsworth.
10 Ashby W.R. (1956) *Introduction to Cybernetics*, Chapman & Hall, London. http://pespmc1.vub.ac.be/books/IntroCyb.pdf (accessed 8 April 2011).
11 Wiener, N. (1948) *Cybernetics: Control and Communications in the Animal and the Machines*, The MIT Press, Cambridge.
12 Ideker, T., Galitski, T., Hood, L. (2001) A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.*, **2**, 343–372.
13 Wolkenhauer, O. (2001) Systems biology: the reincarnation of systems theory applied in biology? *Brief. Bioinform.*, **2** (3), 258–270.
14 Mesarović, M.D. (1968) System Theory and Biology – View of a Theoretician, in: Mesarović, M.D. (Ed.) *Systems Theory and Biology*, Springer, New York, pp. 59–87.
15 Wolkenhauer, O., Hofmeyr, J.-H.S. (2007) An abstract cell model that describes the self-organization of cell function in living systems. *J. Theor. Biol.*, **246** (3), 461–476.
16 Westerhoff, H.V., Palsson, B.O. (2004) The evolution of molecular biology into systems biology. *Nat. Biotechnol.*, **22** (10), 1249–1252.
17 Pittendrigh, C.S. (1958) Adaptation, natural selection, and behavior, in: Roe, A., Simpson, G.G. (Eds) *Behavior and Evolution*, Yale University Press, New Haven, pp. 390–416.
18 Baker, M.D., Wolanin, P.M., Stock, J.B. (2006) Signal transduction in bacterial chemotaxis. *BioEssays*, **28** (1), 9–22.
19 Levine, M., Tjian, R. (2003) Transcription regulation and animal diversity. *Nature*, **424**, 147–151.
20 Huang, S. (2004) Back to the biology in systems biology: what can we learn from biomolecular networks? *Brief. Funct. Genomic. Proteomic.*, **2** (4), 279–297.
21 Dufva, M. (Ed.) (2009) *DNA Microarrays for Biomedical Research*, Methods in Molecular Biology, Vol. 529, Humana Press.
22 Hawkins, R.D., Hon, G.C., Ren, B. (2010) Next-generation genomics: an integrative approach. *Nat. Rev. Genet.*, **11** (7), 476–486.
23 Fields, S., Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature*, **340** (6230), 245–246.
24 Johnsson, N., Varshavsky, A. (1994) Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl Acad. Sci. USA*, **91** (22), 10340–10344.
25 Spiller, D.G., Wood, C.D., Rand, D.A., White, M.R.H. (2010) Measurement of single-cell dynamics. *Nature*, **465** (7299), 736–745.
26 Demirel, Y., Sandler, S.I. (2002) Thermodynamics and bioenergetics. *Biophys. Chem.*, **97**, 87–111.
27 Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W. (1999) From molecular to modular cell biology. *Nature*, **420**, C47–C52.
28 Malygin, A.A., Parakhnevitch, N.M., Ivanov, A.V., Eperon, I.C., Karpova, G.G. (2007) Human ribosomal protein s13 regulates expression of its own gene at the splicing step by a feedback mechanism. *Nucleic Acids Res.*, **35** (19), 6414–6423.
29 Vicsek, T. (2002) Complexity: the bigger picture. *Nature*, **418**, 131.
30 Carlson, J.M., Doyle, J. (2002) Complexity and robustness. *Proc. Natl Acad. Sci. USA*, **99** (Suppl. 1), 2538–2545.
31 Grassberger, P. (1986) Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys.*, **25**, 907–928.

32 Adami, C. (2002) What is complexity? *BioEssays*, **24**, 1085–1094.

33 Barabási, A.-L., Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Gen.*, **5**, 101–114.

34 Kitano, H. (2002) Computational systems biology. *Nature*, **420**, 206–210.

35 Haken, H. (1987) *Advanced Synergetics*, Springer, Berlin.

36 Bak, P., Tang, C., Wiesenfeld, K. (1987) Self-organized criticality: an explanation of the 1/f noise. *Phys. Rev. Lett.*, **59**, 381–384.

37 Carlson, J.M., Doyle, J. (1999) Highly optimized tolerance: a mechanism for power laws in designed systems. *Phys. Rev. E*, **60** (2), 1412–1427.

38 Csete, M.E., Doyle, J.C. (2002) Reverse engineering of biological complexity. *Science*, **295**, 1664–1669.

39 Stelling, J., Sauer, U., Szallasi, Z., Doyle, F.J., Doyle, J. (2004) Robustness of cellular functions. *Cell*, **118**, 675–685.

40 Kitano, H. (2007) Towards a theory of biological robustness. *Mol. Syst. Biol.*, **3**, 137.

41 Lauffenburger, D.A. (2000) Cell signaling pathways as control modules: complexity for simplicity? *Proc. Natl Acad. Sci. USA*, **97** (10), 5031–5033.

42 Gu, X. (2003) Evolution of duplicate genes versus genetic robustness against null mutations. *Trends Genet.*, **19**, 354–356.

43 Tyson, J.J., Chen, K.C., Novak, B. (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.*, **15**, 221–231.

44 Novak, B., Tyson, J.J., Gyorffy, B., Csikasz-Nagy, A. (2007) Irreversible cell cycle transitions are due to systems-level feedback. *Nat. Cell Biol.*, **9** (7), 724–728.

45 Lpez-Avils, S., Kapuy, O., Novk, B., Uhlmann, F. (2009) Irreversibility of mitotic exit is the consequence of systems-level feedback. *Nature*, **459** (7246), 592–595.

46 Clarke, D.J., Giménez-Abián, J.F. (2000) Checkpoints controlling mitosis. *BioEssays*, **22** (4), 351–363.

47 Kearns, W.G., Liu, J.M. (2001) Cell cycle checkpoint genes and aneuploidy: a short review. *Curr. Genomics*, **2** (2), 171–180.

48 Fulka, J., Tesařík, J., Loi, P., Moor, R.M. (2000) Manipulating the human embryo: cell cycle checkpoint controls. *Cloning*, **2** (1), 1–7.

49 Barkai, N., Leibler, S. (1997) Robustness in simple biochemical networks. *Nature*, **387**, 913–917.

50 Alon, U., Surette, M.G., Barkai, N., Leibler, S. (1999) Robustness in bacterial chemotaxis. *Nature*, **397**, 168–171.

51 Hartman, J.L., Garvik, B., Hartwell, L. (2001) Principles for the buffering of genetic variation. *Science*, **291**, 1001–1004.

52 Gell, M., Noort, V., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Khner, S., Rode, M., Suyama, M., Schmidt, S., Gavin, A.-C., Bork, P., Serrano, L. (2009) Transcriptome complexity in a genome-reduced bacterium. *Science*, **326** (5957), 1268–1271.

53 von Dassow, G., Meir, E., Munro, E.M., Odell, G. (2000) The segment polarity network is a robust developmental module. *Nature*, **406**, 188–192.

54 Hanahan, D., Weinberg, R.A. (2000) The hallmarks of cancer. *Cell*, **100** (1), 57–70.

55 Yi, T.-M., Huang, Y., Simon, M.I., Doyle, J.C. (2004) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl Acad. Sci. USA*, **97** (9), 4649–4653.

56 Guisbert, E., Yura, T., Rhodius, V.A., Gross, C.A. (2008) Convergence of molecular, modeling, and systems approaches for an understanding of the *Escherichia coli* heat shock response. *Microbiol. Mol. Biol. Rev.*, **72** (3), 545–554.

57 Kitano, H. (Ed.) (2001) *Foundations of Systems Biology*, The MIT Press, Cambridge, pp. 1–36.

58 Angeli, D., Ferrell, J.E., Sontag, E.D. (2004) Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl Acad. Sci. USA*, **101**, 1822–1827.

59 Becker, V., Schilling, M., Bachmann, J., Baumann, U., Raue, A., Maiwald, T., Timmer, J., Klingmller, U. (2010) Covering

a broad dynamic range: information processing at the erythropoietin receptor. *Science*, **328** (5984), 1404–1408.

60 Freeman, M. (2000) Feedback control of intercellular signalling in development. *Nature*, **408**, 313–319.

61 Kitano, H. (2004) Cancer as a robust system: implications for anticancer therapy. *Nat. Rev.*, **4**, 227–235.

62 Carter, G.W., Rush, C.G., Uygun, F., Sakhanenko, N.A., Galas, D.J., Galitski, T. (2010) A systems-biology approach to modular genetic complexity. *Chaos*, **20** (2), 026102.

63 Guimerà, R., Arenas, A., Díaz-Guilera, A. (2001) Communication and optimal hierarchical networks. *Physica A*, **299**, 247–252.

64 Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J.C., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H. (2002) A genomic regulatory network for development. *Science*, **295**, 1669–1678.

65 Laub, M.T., McAdams, H.H., Feldblum, T., Fraser, C.M., Shapiro, L. (2000) Global analysis of the genetic network controlling a bacterial cell cycle. *Science*, **290**, 2144–2148.

66 Pearl, J. (2000) *Causality*, Cambridge University Press.

67 Dhar, P.K., Zhu, H., Mishra, S.K. (2004) Computational approach to systems biology: from fraction to integration and beyond. *IEEE Trans. Nanobiosci.*, **3** (3), 144–152.

68 Crampin, E.J., Schnell, S., McSharry, P.E. (2004) Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. *Prog. Biophys. Mol. Biol.*, **86** (1), 77–112.

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

70 Endy, D., Brent, R. (2001) Modelling cellular behavior. *Nature*, **409**, 391–395.

71 McAdams, H., Shapiro, L. (1995) Circuit simulations of genetic networks. *Science*, **269**, 650–656.

72 Christopher, R., Dhiman, A., Fox, J., Gendelman, R., Haberichter, T., Kagle, D., Spizz, G., Khalil, I.G., Hill, C. (2004) Data-driven computer simulation of human cancer cell. *Ann. N. Y. Acad. Sci.*, **1020**, 132–153.

73 Müller, F.-J., Laurent, L.C., Kostka, D., Ulitsky, I., Williams, R., Lu, C., Park, I.-H., Rao, M.S., Shamir, R., Schwartz, P.H., Schmidt, N.O., Loring, J.F. (2008) Regulatory networks define phenotypic classes of human stem cell lines. *Nature*, **455** (7211), 401–405.

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

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

76 Elowitz, M.B., Surette, M.G., Wolf, P.-E., Stock, J.B., Leibler, S.J. (1999) Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.*, **181**, 197–203.

77 Schnell, S., Turner, T.E. (2004) Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Prog. Biophys. Mol. Biol.*, **85**, 235–260.

78 Howard, M., Rutenberg, A.D. (2003) Pattern formation inside bacteria: fluctuations due to the low copy number of proteins. *Phys. Rev. Lett.*, **90** (1–4), 128102.

79 Busch, H., Camacho, D., Rogon, Z., Breuhahn, K., Angel, P., Eils, R., Szabowski, A. (2008) Gene network dynamics controlling keratinocyte migration. *Mol. Syst. Biol.*, **4**, 199.

80 Chikofsky, E.J., Cross, J.H. (1990) Reverse engineering and design recovery: A taxonomy. *IEEE Softw.*, **7**, 13–17.

81 He, F., Balling, R., Zeng, A.-P. (2009) Reverse engineering and verification of gene networks: principles, assumptions, and limitations of present methods and future perspectives. *J. Biotechnol.*, **144** (3), 190–203.

82 Strogatz, S.H. (2001) Exploring complex networks. *Nature*, **410**, 268–276.

83 Sharom, J.R., Bellows, D.S., Tyers, M. (2004) From large networks to small

molecules. *Curr. Opin. Chem. Biol.*, **8**, 81–90.
84 Alm, E., Arkin, A.P. (2003) Biological networks. *Curr. Opin. Struct. Biol.*, **13**, 193–202.
85 Barabási, A.-L., Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.
86 Keller, E.F. (2005) Revisiting "scale-free" networks. *BioEssays*, **27** (10), 1060–1068.
87 Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U. (2002) Network motifs: Simple building blocks of complex networks. *Science*, **298**, 824–827.
88 Yan, K.K., Fang, G., Bhardwaj, N., R.P., Alexander, Gerstein, M. (2010) Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *Proc. Natl Acad. Sci. USA*, **107** (20), 9186–9191.
89 Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.-B., Volkert, T.L., Fraenkel, E., Gifford, D.K., Young, R.A. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, **298**, 799–804.
90 Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z.N., Barabási, A.-L. (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*, **427**, 839–843.
91 Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, N., Barabási, A.-L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, **297**, 1551–1555.
92 Pawson, T., Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science*, **300**, 445–452.
93 de la Fuente, A. (2010) From differential expression to differential networking identification of dysfunctional regulatory networks in diseases. *Trends Genet.*, **26**, 326–333.
94 Goh, K.-I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., Barabsi, A.-L. (2007) The human disease network. *Proc. Natl Acad. Sci. USA*, **104** (21), 8685–8690.
95 Saez-Rodriguez, J., Alexopoulos, L.G., Epperlein, J., Samaga, R., Lauffenburger, D.A., Klamt, S., Sorger, P.K. (2009) Discrete logic modeling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.*, **5**, 331.
96 Jong, H. (2002) Modeling and simulation of generic regulatory systems: a literature review. *J. Comput. Biol.*, **9** (1), 67–103.
97 Faur, A., Thieffry, D. (2009) Logical modelling of cell cycle control in eukaryotes: a comparative study. *Mol. Biosyst.*, **5** (12), 1569–1581.
98 Samaga, R., Saez-Rodriguez, J., Alexopoulos, L.G., Sorger, P.K., Klamt, S. (2009) The logic of egfr/erbb signaling: theoretical properties and analysis of high-throughput data. *PLoS Comput. Biol.*, **5** (8), e1000438.
99 Schlatter, R., Schmich, K., Vizcarra, I.A., Scheurich, P., Sauter, T., Borner, C., Ederer, M., Merfort, I., Sawodny, O. (2009) On/off and beyond – a Boolean model of apoptosis. *PLoS Comput. Biol.*, **5** (12), e1000595.
100 Murray, J.D. (1993) *Mathematical Biology*, Springer, Heidelberg.
101 Cross, M.C., Hohenberg, P.C. (1993) Pattern formation outside of equilibrium. *Rev. Mod. Phys.*, **65** (3), 851–1112.
102 Moles, C.G., Mendes, P., Banga, J.R. (2003) Parameter estimation in biochemical pathways: A comparison of global optimization methods. *Genome Res.*, **13**, 2467–2474.
103 Hengl, S., Kreutz, C., Timmer, J., Maiwald, T. (2007) Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics*, **23**, 2612–2618.
104 Bandara, S., Schlder, J.P., Eils, R., Bock, H.G., Meyer, T. (2009) Optimal experimental design for parameter estimation of a cell signaling model. *PLoS Comput. Biol.*, **5** (11), e1000558.
105 Skanda, D., Lebiedz, D. (2010) An optimal experimental design approach to model discrimination in dynamic biochemical systems. *Bioinformatics*, **26** (7), 939–945.
106 Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P. (1993) *Numerical Recipes in C*, 2nd edn, Cambridge University Press, New York.
107 Mendes, P. (1993) Gepasi: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *CABIOS*, **9** (5), 563–571.
108 Goryanin, I., Hodgman, T.C., Selkov, E. (1999) Mathematical simulation and

analysis of cellular metabolism and regulation. *Bioinformatics*, **15** (9), 749–758.

109 Funahashi, A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N., Kitano, H. (2008) Celldesigner 3.5: a versatile modeling tool for biochemical networks. *Proc. IEEE*, **96** (8), 1254–1265.

110 Schmidt, H., Jirstrand, M. (2006) Systems biology toolbox for matlab: a computational platform for research in systems biology. *Bioinformatics*, **22** (4), 514–515.

111 Maiwald, T., Timmer, J. (2008) Dynamical modeling and multi-experiment fitting with Potters wheel. *Bioinformatics*, **24** (18), 2037–2043.

112 Huang, C.Y., Ferrell, J.E. (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl Acad. Sci. USA*, **93** (19), 10078–10083.

113 Turing, A.M. (1952) The chemical basis of morphogenesis. *Phil. Trans. R. Soc. London B*, **237**, 37–72.

114 Meinhardt, H. (2008) Models of biological pattern formation: from elementary steps to the organization of embryonic axes. *Curr. Top. Dev. Biol.*, **81**, 1–63.

115 Kauffman, S.A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, New York.

116 Volz, D., Eigel, M., Athale, C., Bastian, P., Hermann, H., Kappel, C., Eils, R. (2005) *Spatial Modeling and Simulation of Diffusion in Nuclei of Living Cells*, CMSB 2004, Lecture Notes in Computer Science, Springer, Heidelberg, pp. 161–171.

117 Kholodenko, B.N. (2003) Four-dimensional organization of protein kinase signaling cascades: the roles of diffusion, endocytosis and molecular motors. *J. Exp. Biol.*, **206**, 2073–2082.

118 Slepchenko, B.M., Schaff, J.C., Carson, J.H., Loew, L.M. (2002) Computational cell biology: Spatiotemporal simulation of cellular events. *Ann. Rev. Biophys. Biomol. Struct.*, **31**, 423–441.

119 van Kampen, N.G. (1992) *Stochastic Processes in Physics and Chemistry*, Elsevier Science Publishers B.V., Amsterdam.

120 Kaern, M., Elston, T.C., Blake, W.J., Collins, J.J. (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.*, **6** (6), 451–464.

121 Golding, I., Paulsson, J., Zawilski, S.M., Cox, E.C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell*, **123** (6), 1025–1036.

122 Elowitz, M., Levine, A., Siggla, W., Swain, P. (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183–1186.

123 Arkin, A., Ross, J., McAdams, H.H. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, **149** (4), 1633–1648.

124 Huang, S., Ingber, D.E. (2007) A non-genetic basis for cancer progression and metastasis: self-organizing attractors in cell regulatory networks. *Breast Dis.*, **26**, 27–54.

125 Huang, S., Eichler, G., Bar-Yam, Y., Ingber, D.E. (2005) Cell fates as high dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.*, **94** (12), 128701.

126 Huang, S. (2009) Reprogramming cell fates: reconciling rarity with robustness. *BioEssays*, **31** (5), 546–560.

127 Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., Hutchison, C.M. (1999) Ecell: software environment for whole cell simulation. *Bioinformatics*, **15** (5), 72–74.

128 Slepchenko, B.M., Schaff, J.C., Macara, I.G., Loew, L.M. (2003) Quantitative cell biology with the virtual cell. *Trends Cell Biol.*, **13**, 570–576.

129 Dhar, P., Meng, T.C., Somani, S., Ye, L., Sairam, A., Chitre, M., Hao, Z., Sakharkar, K. (2004) Cellware – multi-algorithmic software for computational systems biology. *Bioinformatics*, **20**, 1319–1321.

130 Ander, M., Beltrao, P., Ventura, B.D., Ferkinghoff-Borg, J., Foglierini, M., Kaplan, A., Lemerle, C., Tomas-Oliveira, I., Serrano, L. (2004) SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. *Syst. Biol.*, **1**, 129–138.

131 Kitano, H., Funahashi, A., Matsuoka, Y., Oda, K. (2005) Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.*, **23** (8), 961–966.

132 Novre, N.L., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M.I., Wimalaratne, S.M., Bergman, F.T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y.,

Villger, A., Boyd, S.E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, T.C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D.B., Sander, C., Sauro, H., Snoep, J.L., Kohn, K., Kitano, H. (2009) The systems biology graphical notation. *Nat. Biotechnol.*, **27** (8), 735–741.

133 Lloyd, C.M., Halstead, M.D.B., Nielsen, P.F. (2004) CellML: its future, present and past. *Prog. Biophys. Mol. Biol.*, **85**, 433–450.

134 BioPAX Homepage Internet (2010), http://www.biopax.org/ (accessed 8 April 2011).

135 Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I.I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.-H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Le Novre, N., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

136 Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J., Kitano, H. (2002) The erato systems biology workbench: enabling interaction and exchange between software tools for computational biology. *Pac. Symp. Biocomput.*, 450–461.

137 Gibson, M.A., Bruck, J. (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.*, **104**, 1876–1889.

138 Gillespie, D.T. (2001) Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, **115**, 1716–1733.

139 Morton-Firth, C.J. (1998) Stochastic simulation of cell signalling pathways. PhD thesis, University of Cambridge.

140 Kierzek, A.M. (2002) Stocks: stochastic kinetic simulations of biochemical systems with Gillespie algorithm. *Bioinformatics*, **18** (3), 470–481.

141 Lindvall, J.M., Emelie, K., Blomberg, M., Smith, C.I.E. (2003) In silico tools for signal transduction research. *Brief. Bioinform.*, **4**, 315–324.

142 Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., Mering, C. (2009) String 8 – a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **37** (Database issue), D412–D416.

143 Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., Hirakawa, M. (2010) Kegg for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38** (Database issue), D355–D360.

144 Arkin, A.P. (2001) Synthetic cell biology. *Curr. Opin. Biotechnol.*, **12** (6), 638–644.

145 Swedlow, J.R., Goldberg, I., Brauner, E., Sorger, P.K. (2003) Informatics and quantitative analysis in biological imaging. *Science*, **300**, 100–102.

146 Goldberg, I., Allan, C., Burel, J.-M., Creager, D., Falconi, A., Hochheiser, H., Johnston, J., Mellen, J., Sorger, P.K., Swedlow, J.R. (2005) The Open Microscopy Environment (OME) Data Model and XML File: Open Tools for Informatics and Quantitative Analysis in Biological Imaging. *Genome Biol.*, **6**, R47.

147 Searls, D.B. (2002) The language of the genes. *Nature*, **420**, 211–217.

148 François, P., Hakim, V. (2004) Design of genetic networks with specified functions by evolution in silico. *Proc. Natl Acad. Sci. USA*, **101** (2), 580–585.

149 Butcher, E.C., Berg, E.L., Kunkel, E.J. (2004) Systems biology in drug discovery. *Nat. Biotechnol.*, **22** (10), 1253–1259.

150 Hood, L. (2008) A personal journey of discovery: developing technology and changing biology. *Annu. Rev. Anal. Chem. (Palo Alto CA)*, **1**, 1–43.

151 Fitzgerald, J.B., Schoeberl, B., Nielsen, U.B., Sorger, P.K. (2006) Systems biology and combination therapy in the quest for clinical efficacy. *Nat. Chem. Biol.*, **2** (9), 458–466.

152 Schoeberl, B., Pace, E.A., Fitzgerald, J.B., Harms, B.D., Xu, L., Nie, L., Linggi, B., Kalra, A., Paragas, V., Bukhalid, R.,

Grantcharova, V., Kohli, N., West, K.A., Leszczyniecka, M., Feldhaus, M.J., Kudla, A.J., Nielsen, U.B. (2009) Therapeutically targeting erbb3: a key node in ligand-induced activation of the erbb receptor-pi3k axis. *Sci. Signal.*, **2** (77), ra31.

153 Hendriks, B.S. (2010) Functional pathway pharmacology: chemical tools, pathway knowledge and mechanistic model-based interpretation of experimental data. *Curr. Opin. Chem. Biol.*, **14** (4), 489–497.

154 Purnick, P.E.M., Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell. Biol.*, **10** (6), 410–422.

155 Kiel, C., Yus, E., Serrano, L. (2010) Engineering signal transduction pathways. *Cell*, **140** (1), 33–47.

156 Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E.A., Young, L., Qi, Z.-Q., Segall-Shapiro, T.H., Calvey, C.H., Parmar, P.P., Hutchison, C.A., Smith, H.O., Venter, J.C. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329** (5987), 52–56.

157 Hasty, J., McMillen, D., Collins, J.J. (2002) Engineered gene circuits. *Nature*, **420**, 224–230.

158 Guet, C.C., Elowitz, M.B., Hsing, W., Leibler, S. (2002) Combinatorial synthesis of genetic networks. *Science*, **296**, 1466–1470.

159 May, R.M. (2004) Uses and abuses of mathematics in biology. *Science*, **303**, 790–793.

160 Vilar, J.M.G., Kueh, H.Y., Barkai, N., Leibler, S. (2002) Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl Acad. Sci. USA*, **99**, 5988–5992.