In silico multicellular systems biology and minimal genomes

Eric Werner

The *in vivo* and *in silico* understanding of genomes and networks in cellular and multicellular systems is essential for drug discovery for multicellular diseases. *In silico* methodologies, when integrated with *in vivo* engineering methods, lay the groundwork for understanding multicellular organisms and their genomes. The quest to construct a minimal cell can be followed by designed, minimal multicellular organisms. *In silico* multicellular systems biology will be essential in the design and construction of minimal genomes for minimal multicellular organisms. Advanced methodologies come to light that can aid drug discovery. These novel approaches include multicellular pharmacodynamics and networked multicellular pharmacodynamics.

Eric Werner
Cellnomica
PO Box 1422
Fort Myers, FL 33928-1422, USA
e-mail:
eric.werner@cellnomica.com

▼ Drug discovery for multicellular diseases such as cancer can be enhanced by understanding the interaction between potential drug candidates, genomic networks, signaling networks, metabolic networks and multicellular processes. A system level analysis can weed out possibly dangerous drug interactions with a multicellular system, as well as point to potential targets for drug design. Because of the complexity of multicellular systems, *in silico* multicellular modeling and simulation will become an essential component in the drug discovery process. The effort to create minimal genomes for minimal cells will provide the prerequisites for the *in silico* design and *in vivo* engineering of minimal multicellular genomes (mMCGs) for minimal multicellular systems (mMCSs). This will result in testable *in silico* multicellular modeling and simulation that is necessary for drug target discovery, drug discovery and design.

Despite an ever-increasing inundation of genomic data [1,2], we still do not have a solid understanding of how genomes relate to phenotype, nor do we understand many of the processes within cells. We do not even know why it is that the genes of humans and chimps are virtually identical and yet they differ significantly in morphology and behavior. How is this possible? For those who think that chimpanzees are not much different from ourselves, consider the worm or the fruit fly that also share many of their genes with us. The only answer must be that the difference in structure and complexity resides not so much in the genes, but rather in the regulation of those genes. And, regulation occurs in the noncoding regions of genomes. Indeed, this is becoming the accepted view among developmental biologists [3–5]. The focus then shifts from genes to whole genomes and how they are organized. Let us call the organization of a genome the 'regulatory architecture' or 'genome architecture'. The regulatory architecture is the global network in the genome that regulates the dynamics and the development of an organism.

Genome semantics or how does a chimpanzee differ from a human?

The genes of a chimpanzee and a human are virtually identical, therefore, it must be the regulatory genome architecture where the genomes and humans and chimpanzees will differ sufficiently to account for the differences in development, morphology and behavior.

The central problem of post-genomic biology and medicine is to understand the meaning of genomes. To understand genomes we need to view them in their biological context. This includes the context of the cell and the dynamic context of a developing multicellular organism. Traditionally, this is done in the laboratory. A complementary approach is to model and simulate the functioning of genomes using *in silico* models of cells and multicellular organisms. Here, a hybrid, multilevel approach is propsed, which combines *in silico* and *in vivo* methods.

A genome requires the cell to interpret what its genome means. In the case of the genome,

1359-6446/03/$ – see front matter PII: S1359-6446(03)02918-0

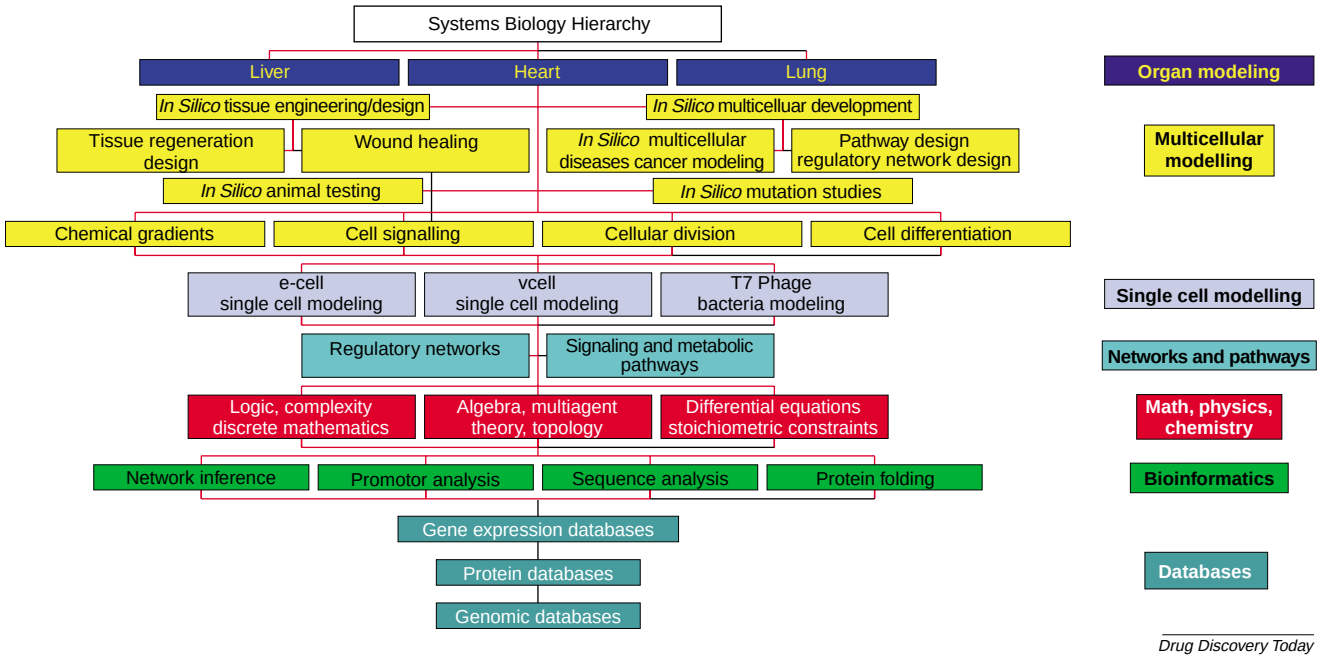

Figure 1. The systems biology hierarchy is a rough working chart of a broad spectrum of evolving concepts and research. There are several possible interpretations: (1) It is a modeling hierarchy with higher level models combining and abstracting concepts and models from lower levels. (2) It is a research hierarchy where levels correspond to areas of research in systems biology with higher levels depending more or less on lower levels. (3) It is an informational–ontological hierarchy that is modular or compositional. Many of the levels correspond to levels of information and ontology with information and objects at higher levels being constructed using lower level information and objects. Information flows up and down these levels. (4) It is an organizational hierarchy; as we go up the hierarchy we get more encompassing organization. (5) It is a semantic hierarchy where the genome is interpreted differently at each level. Note, mathematics, physics and chemistry are at the center radiating out to higher and lower levels. The lower levels are more data intensive.

however, this is more complicated because part, if not all, of the interpreter (object semantics [6,7]) within the cell is generated by the genome itself. Thus, the genome contains within itself its own interpretation. We will see that genome semantics must deal with a hierarchy of levels of information and ontology. There are two related hierarchies: a modeling hierarchy and a semantic hierarchy. We first describe the modeling hierarchy in systems biology.

The systems biology hierarchy

Computational systems biology [8–14] can be described as a set of areas of research and modeling that fall into a hierarchy (Figure 1). In this hierarchy, all levels have an equal contributing value; for example, while the higher levels tend to be supported by the lower levels, a higher level might also be used to organize and give meaning to a lower level. To illustrate this, mutations in the genome, at a low level, are described in terms of their effects on the morphology at the level of the organism, a high level of information and ontology. This also shows that a low level can have a direct effect on a much higher level of information.

There are different levels of detail and scope in modeling biological systems. Each level has its own laws but the levels do interact. At the lowest level, we have traditional bioinformatics dealing with genome and protein databases. At the next level is software that analyzes and makes predictions based on data. This includes promoter analysis [15,16], network inference [17,18] and the prediction of how proteins fold [19,20]. The next level is really the basis for all the levels – theoretical computer science, mathematics, physics and chemistry. There are many approaches to the analysis of genomic regulatory networks, including logic and differential equations [21–23] and constraint-based modeling [24]. Multiagent systems theory is relevant when we consider complex systems of interacting agents [6,25]. These agents can be at different levels of ontology; for example, proteins might be considered as agents but cells might also be modeled as agents. The level consisting of regulatory networks and pathways provides the logical glue that controls the systems at the higher levels [23]. At the next level, single cell models have and are being developed where pathways and networks are the dominant components that are being modeled [26–28].

Multicellular modeling goes beyond single cells. It is concerned with the interactions between cells, including cell signaling, growth, division and differentiation. At the level of multicellular simulation, a whole new set of phenomena appear and are amenable to study that are left out of the lower levels. Some of these include tissue regeneration, wound healing, organ regeneration, tissue engineering, modeling and simulating the dynamics of multicellular diseases, the effect of pathways and networks on multicellular development and disease. In addition, the effects of *in silico* mutations on the dynamics and development of the multicellular system in question can be studied.

At a higher level, organs and the relationships between organs are modeled. These models abstract away from individual cell structure to more generic properties [29].

Properties of systems

Ideally, in an analogous way to the relationship between phenomenological and statistical thermodynamics, systems biology will permit the modeling of properties at different levels of abstraction and ontology. The information required depends on the level of information that we are modeling. Each level of information has its own ontology of objects and relations. Thus, we can have models that are incomplete and yet accurate at a given level. For example, one might model the effects of homeotic mutations without necessarily modeling all the metabolic pathways in the cell.

The function or meaning of a part depends on the context of the other parts with which it interacts. For example, the same gear in a mechanical clock might be part of a component to move the second hand or the hour hand. Analogously, the function and meaning of a signal will depend on the multicellular context and not just the genome that generated the signal.

The semantic hierarchy and the meaning of genomes

To the hierarchy of levels modeling in systems biology, there corresponds a semantic hierarchy of levels of information and ontology. Furthermore, there is an information flow between these levels. The specification of the semantics of genomes requires that we understand how the cell interprets the genome in different contexts and levels. At the lowest level, we have the transcription into RNA and then translation of mRNA into proteins. At a higher level, proteins have particular functions within the cell. Through this we can assign a meaning to the gene that generated the protein (not considering alternate splicing for the moment). At a still higher level the protein and its correlate gene might have functions in the cells external behavior, such as movement, sending a signal or in the interpretation of incoming signals.

At a further higher level of information, a signal might be involved in the dynamics of development, for example, in the formation a multicellular boundary. At a higher level still, a gene might get its meaning from the physiological consequences of mutating that gene. For example, a single point mutation in the fruit fly can change a normal fly into one that has four wings [30,31]; developmental biology is essential to the understanding of multicellular systems [32,33].

Thus, to fully understand a genome we need to understand it at different levels of abstraction, information and ontology. In addition, as the example of the four-winged fruit fly shows, a change at the lowest level can have a direct effect on the structure of the highest levels, namely the development and resulting anatomy of the organism.

We are led to the conclusion that to understand genomes we need to understand their semantics in the context of the cell, as well as the context of the developing multicellular organism that it generates. Thus, the meaning of genomes is resolved into a hierarchy of levels of meaning, which we call the genome semantic hierarchy.

Problems with the bottom-up approach

Although there is a lot of data being generated in laboratories about the expression of genes, RNA and proteins, we can not interpret the data unless we have a semantics of the given genome. In other words, we need to understand how the genome functions in the multicellular system. How do we get the semantics? The dominant approach is bottom-up; we create mutations and observe the effects. However, genomes and their organisms are highly complex and, as a result, using a direct bottom-up search for the meaning of genomes is computationally unfeasible. In computer science, when the search space is NP-complete it means that the search space is so vast that no computer can search through it in any realistic time. Depending on the size of the search space, an NP-complete (nondeterministic polynomial time complete) problem could take hundreds of years find an optimal solution. In addition, the search space increases exponentially, so to deal with this unmanageable situation, computer scientists are forced to complement bottom-up methods with top-down methods, including heuristic information to reduce the search space. For example, chess players use heuristic information, based on their intuition and experience to be able to compete with computers. Thus, to understand genomes we need a multilevel approach that models many levels of information.

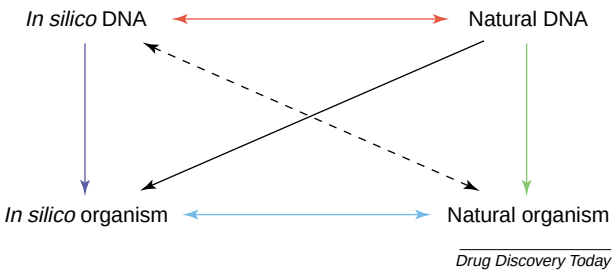

Figure 2. Relationship between *in silico* and natural MCOs. The red arrow is the translation relationship between the *in silico* and natural genomes. The blue arrow on the left represents the *in silico* generation relation from genome to multicellular organism. The green arrow on the right represents the generation relation between the natural genome and the *in vivo* or *in vitro* organism. The bottom light blue arrow is the correspondence relation between the generated organisms. The middle black solid arrow points to the possibility via the translation relation of predicting the phenotype of a natural genome *in silico* without generating it *in vivo*. The middle dashed black arrow points to the possibility of designing and/or reverse engineering natural organisms by way of the translation relation.

In silico genomes

One way of reducing the search space for the semantics of genomes is to conduct *in silico* simulations of genomes that lead to multicellular phenomena that correspond to natural multicellular phenomena. In such an approach, the semantics of genomes is constrained by what we know from research in molecular biology, cell biology and genetics, as well as a century of experimentation in developmental biology. Modeling multicellular or cellular processes is, in essence, a process of theory construction where computer software is used to develop and test theories of the function of genomes, their architecture and their semantics. This is a fundamentally different approach than that of going directly from data to models.

Minimal genomes and minimal cells

Complementary to a virtual, *in silico* approach that models the semantics of single cell and multicellular genomes, is the *in vivo* construction of genomic networks that regulate cell processes. The work by Venter and others to construct a minimal cell with a minimal genome is part of this effort [34–39]. Whereas Venter's group is trying to construct a minimal cell bottom-up, others [37–40] are reducing simple bacterial genomes to find a minimal set of essential genes. The idea is to use forward engineering to construct an artificial organism. The method presupposes a group of components that are more or less understood. These components are then combined to construct further components. Ultimately, this leads to the systematic building of the entire cell. By finding the combination of components that are essential, those components can be systematically modified and added to, to gain new functions.

The starting point is what are known to be almost minimal genomes. By taking the intersection of the two smallest known genomes, their commonalities can be found. This results in an initial set of genes that could be presumed to be minimal. This approach makes the implicit assumption that the common factor of two organisms must be minimal. However, it could be that there is no minimal cell as such. Two single-celled organisms O1 and O2 could be minimal and yet not have a common core. Imagine we have a goal state G that is necessary for survival. Some of the genes might interact cooperatively so that in organism O1, gene A and gene B yield goal state G, yet in the other organism O2, gene A and gene C also yield the same goal state G. In such a case both genomes are minimal in that one can not delete any of their genes without loss of viability, and yet they are not identical. In cases of coevolution, this could have occurred. If organisms really do consist of modular, functional units, then functionally equivalent units yet with distinct components constructed from different, nonhomologous genes might have evolved independently. Hence, it might not be that the intersection of two minimal organisms is really minimal, it might be a nonviable organism. If this is the case, then the above enterprise could be even more difficult to achieve.

An additional but essential difficulty with the construction of the minimal cell is that the regulatory regions that control the functioning of the cell must be sufficiently understood. At present, there is still a great deal of mystery surrounding the functioning of the promoter and other regulatory regions in the control of the cell. Initially, Venter wants to avoid the issue of regulatory regions [35]. Whether and how these important control regions can be sidelined while constructing the minimal cell is still an open question.

Minimal multicellular genomes

If the construction of a minimal cell can be achieved, what happens next? I believe that the next step is to investigate the regulatory properties of minimal cells. After that it becomes possible to investigate minimal multicellular genomes (mMCGs) and their organisms. An mMCG for a multicellular system is the simplest genome that is capable of generating that system. In other words, mMCGs generate mMCSs. An mMCS is a multicellular structure that develops from a single cell using a minimum of cellular structure and genomic information. A minimal multicellular organism (mMCO) is a type of mMCS that is robust [12,41–43], capable of sustaining itself and is usually able

to reproduce. Unlike an mMCO, an mMCS need not contain specialized cell systems required for self sufficiency or self reproduction. The multicellular system is minimal in the sense that the genome necessary for the system to develop is minimal and the cell structure of the differentiated cells is minimal.

Note that mMCSs can be *in vivo* or *in silico*. Indeed, *in vivo* systems might first be designed *in silico* before being translated into actual DNA to construct an *in vivo* genome, which is then inserted into a minimal starter cell. This process is illustrated in Figure 2.

As in the case of minimal single cells, there might be no unique minimal genome; two genomes of equal size might be able to develop a single cell into an equivalent multicellular structure. Because of the enormous complexity of natural genomes and their organisms, a science of minimal genomes for mMCOs will permit us to understand the fundamental genes and regulatory networks that are necessary and sufficient for the development of constructed and natural cells. Given that basis we can then go on to gain a deeper understanding of more complex natural genomes. It would be expected that natural genomes have at least the following additional sources of complexity. First, there will be archaic sub-genomes for functions and structures that are no longer needed for the survival of the modern organism (Werner, E., unpublished; see http://www.cellnomica.com). Second, there will be complexity from genes and networks that are required for the survival of the organism in complex environments. Third, there will be complexity from genes and networks for features unique to a particular type of organism. Finally, there will be differences in individuals genomes that confer the uniqueness to each individual's ontogeny and phenotype.

Just as models of single cells and their networks have an essential role in the construction of the minimal cell, so too *in silico* modeling and simulation of genomes in multicellular development will have a dominant role in the design and construction of mMCGs. In fact, we will reach a point were most of the modeling and design of the genome will first be done in software before the genome is inserted into a starter cell to then develop into the corresponding multicellular structure *in vitro* or *in vivo*.

Engineered mMCOs and drug discovery

This forward engineering process will open up new areas of biotechnology as well as multicellular pharmacodynamics (MCPD; see subsequent sections). In particular, networked MCPD might see great advantages in that an *in silico* model of an mMCO can then be tested and corrected by how the corresponding natural mMCO responds to a drug. One problem with the minimalist approach is that because the cell is minimal it might lack some of the drug responses of the normal cell. However, this is outweighed by the advantages that lead to a precise understanding of the components and the organization of cells and mMCOs and their reaction to a given drug. For, example a drug candidate might be predicted (*in silico)* to affect certain targets in the genomic network of the *in silico* mMCO; however, the corresponding *in vivo* mMCO might or might not respond as predicted. If it does not respond, either the *in silico* model of the mMCO is modified or the *in vivo* genome or the cell is changed. Thus, once we have a working *in silico* mMCO, we can then increase the complexity gradually until it approaches that of natural MCOs.

In summary, it becomes apparent that the project of constructing the minimal cell based on a minimal genome, and the even more important construction of minimal genomes for mMCOs, has important consequences for both biotechnology and the pharmaceutical industry. It could well revolutionize the way we do biology and the life sciences.

Systems biology and drug discovery

As described previously, the genome can and should be interpreted at different levels of information and ontology. In the drug discovery process, there is also a realization by scientists that an ontological view restricted to the molecular level might hinder the drug discovery process. There is a movement to complement the molecular view with a systemic view in drug design [44]. Regulatory networks require a systemic understanding. Although it is certainly true that there are powerful and effective drugs with a broad range of applicability, the complexity of information enables drug design with precise effects on network behavior. However, to achieve this the drug design process needs to include the genomic network and its interpretation by the cell.

For example, in addition to the standard ADMET properties, other higher level properties and objects become relevant including regulation, repression, activation, promoters, enhancers, signaling, gradients, differentiation and temporal dynamics.

Multi-cellular pharmacodynamics

Pharmacokinetics is the study of the distribution of drugs in organs and tissue. Pharmacodynamics (PD) goes a step further and attempts to get at the causes and workings of ADMET properties [44]. We propose to extend PD one step further.

Multicellular modeling of PD uses multicellular models where cells can be in diversely differentiated states. It uses multicellular tissue to model and dynamically simulate and display the effect of drug distribution and other ADMET

properties and is a hybrid approach that combines molecular information with cell and tissue level information. The ADMET properties are modeled in the context of a complex multicellular structure. The multicellular structure can be static, dynamic and even developmentally dynamic, where the tissue or organ in question is developing (growing and differentiating over time).

MCPD is the study of the static and dynamic properties and relationships between a set of drugs and a dynamic and diverse multicellular 4D organization. In contrast to simple compartmental models, MCPD models open the door to fine-grained modeling of tissue, morphology and cellular level response to drug ADMET properties.

Networked multicellular pharmacodynamics

Furthermore, MCPD models can be extended to model regulatory genomic networks together with signal transduction pathways, as part of a complex of interacting components in the cell; these are known as networked-MCPD (Net-MCPD) models. In this way, drug interactions with the cell, the genome, the cell signaling dynamics and the multicellular system can also become accessible to modeling. In this approach, many levels of the systems biology hierarchy are involved in modeling and simulation to the benefit of the drug discovery process.

In silico cancer modeling and simulation

For example, in cancer the regulatory networks in the genome and cell signaling dynamics can have a key role in the etiology, ontogenesis and dynamics of the disease. In this case, drug ADMET properties must be supplemented by additional properties that influence the dynamics of relevant cellular disease states. MCPD models enhanced with genome and cell signaling components might give us a deeper insight into the dynamics of cancers and their response to drugs in a dynamic multicellular context.

In a computer study of *in silico* cancers, the cancers could be prevented at many different sites in the intracellular pathways (Werner, E., unpublished; see http://www.cellnomica.com). However, most of the sites led to severe developmental or metabolic abnormalities. Only a few of the sites had minimal side effects. This *in silico* study has relevance to understanding the *in vivo* situation in real cancers, in that it shows that complex dynamic systems can have different responses to the same drug at different points in the development of the organism. Therefore, the effect of a drug on the dynamics of cell growth might depend on the dynamic state of the cell with which it interacts.

The drug discovery process can be aided by different *in silico* methods corresponding to different levels in the systems biology hierarchy. There are low level methods, such as QSAR and Lipinski's Rule of Five [45,46]. In the future we expect these valuable methods to be supplemented by methods that focus on higher levels of the systems biology hierarchy, including *in silico* PD, *in silico* MCPD and Net-MCPD. In some ways, pharmaceutical research is already at the level of Net-MCPD because recent drug targets aim precisely at intracellular networks and pathways. What is emerging and the promise of *in silico* Net-MCPD, is the ability to computationally test some of these drug interactions with cellular networks in the context of a dynamic multicellular space–time process.

Conclusion

At present, drug discovery is still dominated by a bottom-up approach that mimics the flow of information dictated by the Central Dogma [47]. However, there are inherent limitations with this approach because of the NP-complexity of the search space. Here, an alternative, multileveled approach has been proposed that includes *in silico* multicellular systems biology in tandem with *in vivo* forward and reverse engineering methods to analyze and design minimal genomes for mMCSs. A test bed for drug discovery and development will be provided by mMCOs. Moreover, the higher the level of information that can be modeled, the greater the search space that is accessible to exploration, prediction and reduction. In addition, exploring and reducing the search space is essential for drug discovery and testing and *in silico* multicellular systems biology can help in the understanding of processes involved in wound healing, tissue and organ regeneration.

We showed that multicellular systems biology could provide new methodologies, such as MCPD and Net-MCPD that can aid the drug discovery process. If, for example, a drug has a direct effect on a high level network involved in multicellular development, *in silico* mMCOs could point to possibly serious, even disastrous side effects. By contrast, high-level network analysis, modeling and simulation of *in silico* and *in vivo* mMCOs can also illuminate those targets with the fewest side effects. Hence, as models and simulations become more accurate, they could become indispensable for drug discovery and drug safety. Therefore, multilevel systems analysis *in vivo* combined with *in silico* design, modeling and simulation provides a new foundation for drug discovery.

References

1 Frishman, D. *et al.* (1998) Comprehensive, comprehensible, distributed and intelligent databases: current status. *Bioinformatics* 14, 551–561
2 Frishman, D. *et al.* (2002) Online genomics facilities in the new millennium. *Pharmacogenomics* 3, 265–271
3 Bolouri, H. and Davidson, E.H. (2002) Modeling transcriptional regulatory networks. *Bioessays* 24, 1118–1129

4 Brown, C.T. *et al.* (2002) New computational approaches for analysis of cis-regulatory networks. *Dev. Biol.* 246, 86–102
5 Davidson, E.H. (2001) *Genomic Regulatory Systems: Development and Evolution*, Academic Press
6 Werner, E. (1996) Logical foundations of distributed artificial intelligence. In *Foundations of Distributed Artificial Intelligence*, (O'Hare, G. and Jennings, N., eds), Wiley
7 Werner, E. (1992) The design of multi-agent systems. In *Decentralized AI* (Vol 3) (Werner, E. and Demazeau, Y., eds), Elsevier Science Publishers B.V.
8 Kitano, H. (2002) Systems biology: a brief overview. *Science* 295, 1662–1664
9 Kitano, H. (2002) Computational systems biology. *Nature* 420, 206–210
10 Werner, E. (2003) *In silico* cell signaling underground. *Sci. STKE* 170, PE8
11 Carlson, J.M. and Doyle, J. (2002) Complexity and robustness. *Proc. Natl. Acad. Sci. U. S. A.* 99(Suppl 1), 2538–2545
12 Csete, M.E. and Doyle, J.C. (2002) Reverse engineering of biological complexity. *Science* 295, 1664–1669
13 Palsson, B. (2000) The challenges of *in silico* biology. *Nat. Biotechnol.* 18, 1147–1150
14 Palsson, B.O. (1997) What lies beyond bioinformatics? *Nat. Biotechnol.* 15, 3–4
15 Werner, T. (2001) Cluster analysis and promoter modelling as bioinformatics tools for the identification of target genes from expression array data. *Pharmacogenomics* 2, 25–36
16 Werner, T. (2003) Promoters can contribute to the elucidation of protein function. *Trends Biotechnol.* 21, 9–13
17 van Someren, E.P. *et al.* (2002) Genetic network modeling. *Pharmacogenomics* 3, 507–525
18 Bock, J.R. and Gough, D.A. (2003) Whole-proteome interaction mining. *Bioinformatics* 19, 125–134
19 Godzik, A. (2003) Fold recognition methods. *Methods Biochem. Anal.* 44, 525–546
20 Bartlett, G.J. *et al.* (2003) Inferring protein function from structure. *Methods Biochem. Anal.* 44, 387–407
21 Smolen, P. *et al.* (2000) Modeling transcriptional control in gene networks–methods, recent results, and future directions. *Bull. Math. Biol.* 62, 247–292
22 Smolen, P. *et al.* (2000) Mathematical modeling of gene networks. *Neuron* 26, 567–580
23 Neves, S.R. and Iyengar, R. (2002) Modeling of signaling networks. *Bioessays* 24, 1110–1117
24 Edwards, J.S. *et al.* (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* 19, 125–130
25 Werner, E. (1988) Toward a Theory of Communication and Cooperation for Multiagent Planning. *Theoretical Aspects of Reasoning About Knowledge: Proceedings of the Second Conference*, 129–143
26 Tomita, M. (2001) Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol.* 19, 205–210
27 Tomita, M. *et al.* (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics* 15, 72–84
28 Loew, L.M. and Schaff, J.C. (2001) The Virtual Cell: a software environment for computational cell biology. *Trends Biotechnol.* 19, 401–406
29 Noble, D. (2002) Modelling the heart: insights, failures and progress. *Bioessays* 24, 1155–1163
30 Lewis, E.B. (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565–570
31 Lewis, E.B. (1998) The bithorax complex: the first fifty years. *Int. J. Dev. Biol.* 42, 403–415
32 Nusslein-Volhard, C. and Wieschaus, E. (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795–801
33 St Johnston, D. and Nusslein-Volhard, C. (1992) The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 68, 201–219
34 Venter, C. (2003) Charting the biologic frontier. Interview by Val J. Halamandaris. *Caring* 22, 32–40
35 Venter, C. (2003) Talk on minimal cell and Q&A. *World Drug Discovery and Development Summit*, 17–19 February 2003, Copenhagen, Denmark
36 Schilling, C.H. *et al.* (2002) Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* 184, 4582–4593
37 Kolisnychenko, V. *et al.* (2002) Engineering a reduced *Escherichia coli* genome. *Genome Res.* 12, 640–647
38 Kobayashi, K. *et al.* (2003) Essential Bacillus subtilis genes. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4678–4683
39 Fraser, C.M. *et al.* (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397–403
40 Hutchison, C.A. *et al.* (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* 286, 2165–2169
41 Carlson, J.M. and Doyle, J. (2000) Highly optimized tolerance: robustness and design in complex systems. *Phys. Rev. Lett.* 84, 2529–2532
42 Eldar, A. *et al.* (2002) Robustness of the BMP morphogen gradient in *Drosophila* embryonic patterning. *Nature* 419, 304–308
43 Wilke, C.O. and Adami, C. (2003) Evolution of mutational robustness. *Mutat. Res.* 522, 3–11
44 Jackson, R.C. Predictive software for drug design and development. *Pharm. Dev. Reg.* 1, 159–168
45 Lipinski, C.A. (2003) Chris Lipinski discusses life and chemistry after the Rule of Five. *Drug Discov. Today* 8, 12–16
46 Lipinski, C.A. *et al.* (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26
47 Crick, F. (1970) Central dogma of molecular biology. *Nature* 227, 561–563