The evolution of molecular biology into systems biology

Hans V Westerhoff[1] & Bernhard O Palsson[2]

Systems analysis has historically been performed in many areas of biology, including ecology, developmental biology and immunology. More recently, the genomics revolution has catapulted molecular biology into the realm of systems biology. In unicellular organisms and well-defined cell lines of higher organisms, systems approaches are making definitive strides toward scientific understanding and biotechnological applications. We argue here that two distinct lines of inquiry in molecular biology have converged to form contemporary systems biology.

Whereas the foundations of systems biology-at-large are generally recognized as being as far apart as 19th century whole-organism embryology and network mathematics, there is a school of thought that systems biology of the living cell has its origin in the expansion of molecular biology to genome-wide analyses. From this perspective, the emergence of this 'new' field constitutes a 'paradigm shift' for molecular biology, which ironically has often focused on reductionist thinking. Systems thinking in molecular biology will likely be dominated by formal integrative analysis going forward rather than solely being driven by high-throughput technologies.

It is, however, incorrect to state that integrative thinking is new to molecular biology. The first molecular regulatory circuits were mapped out over 40 years ago. The feedback inhibition of amino acid biosynthetic pathways was discovered in 1957 (refs. 1,2), and the transcriptional regulation associated with the glucose-lactose diauxic shift led to the definition of the *lac* operon and the elucidation of its regulation[3]. With the study of these regulatory mechanisms, admittedly on a small scale, molecular biologists began to apply systems approaches to unravel the molecular components and logic that underlie cellular processes, often in parallel with the characterization of individual macromolecules. High-throughput technologies have made the scale of such inquiries much larger, enabling us to view the genome as the 'system' to study. Thus, the popular contemporary view of systems biology may be synonymous with 'genomic' biology.

This article discusses two historical roots of systems biology in molecular biology (**Fig. 1**). Although we briefly outline the more familiar first root—which stemmed from fundamental discoveries about the nature of genetic material, structural characterization of macromolecules and later developments in recombinant and high-throughput technologies—more emphasis is placed on the second root, which sprung from nonequilibrium thermodynamics theory in the 1940s, the elucidation of biochemical pathways and feedback controls in unicellular organisms and the emerging recognition of networks in biology. We conclude by discussing how these two lines of work are now merging in contemporary systems biology.

Scaling-up molecular biology

In the decades following its foundational discoveries of the structure and information coding of DNA and protein, molecular biology blossomed as a field, with a series of breathtaking discoveries (**Fig. 1**). The description of restriction enzymes and cloning were major breakthroughs in the 1970s, ushering in the era of genetic engineering and biotechnology. In the 1980s, we began to see the scale-up of some of the fundamental experimental approaches of molecular biology. Automated DNA sequencers began to appear and reached genome-scale sequencing in the mid-1990s[4,5]. Automation, miniaturization and multiplexing of various assays led to the generation of additional 'omics' data types[6,7].

The large volumes of data generated by these approaches led to rapid growth in the field of bioinformatics, again largely emanating from the reductionist perspective. Although this effort was mostly focused on statistical models and object classification approaches in the late 1990s, it was recognized that a more formal and mechanistic framework was needed to analyze multiple high-throughput data types systematically[8,9]. This need led to efforts toward genome-scale model building to analyze the systems properties of cellular function.

Molecular self-organization

Even before the first key events in the history of molecular biology, several lines of reasoning revealed that integration of multiple molecular processes is fundamental to the living cell. Biochemical processes necessitate the production of entropy (chaos in the thermodynamic sense) as driving force. The paradox felt by many, but expressed by Schrödinger in his war-time lectures[10], was how one could explain the progressive ordering that occurs in developmental biology (that is, the 'self-organization,' decrease in chaos) when entropy ('chaos') must be increased.

The answer was that one process could produce order (negative entropy or negentropy) provided it was coupled to a second process that produced more chaos (entropy): coupling, another word for integration of processes, is therefore essential for life. Onsager[11] provided the basis for this concept by stressing the significance of the coupling of dissimilar processes. He is also relevant because he discovered a law for such systems of coupled processes: close to equilibrium the dependence of the one process rate on the driving force of the other process should equal the dependence of the other process rate on the

[1]Departments of Molecular Cell Physiology and Mathematical Biochemistry, BioCentrum Amsterdam, De Boelelaan 1085, NL-108, HV Amsterdam, the Netherlands. [2]Department of Bioengineering, University of California-San Diego, 9500 Gilman Drive, La Jolla, California 92093-0412, USA. Correspondence should be addressed to H.V.W. (hw@bio.vu.nl) or B.O.P. (palsson@ucsd.edu).
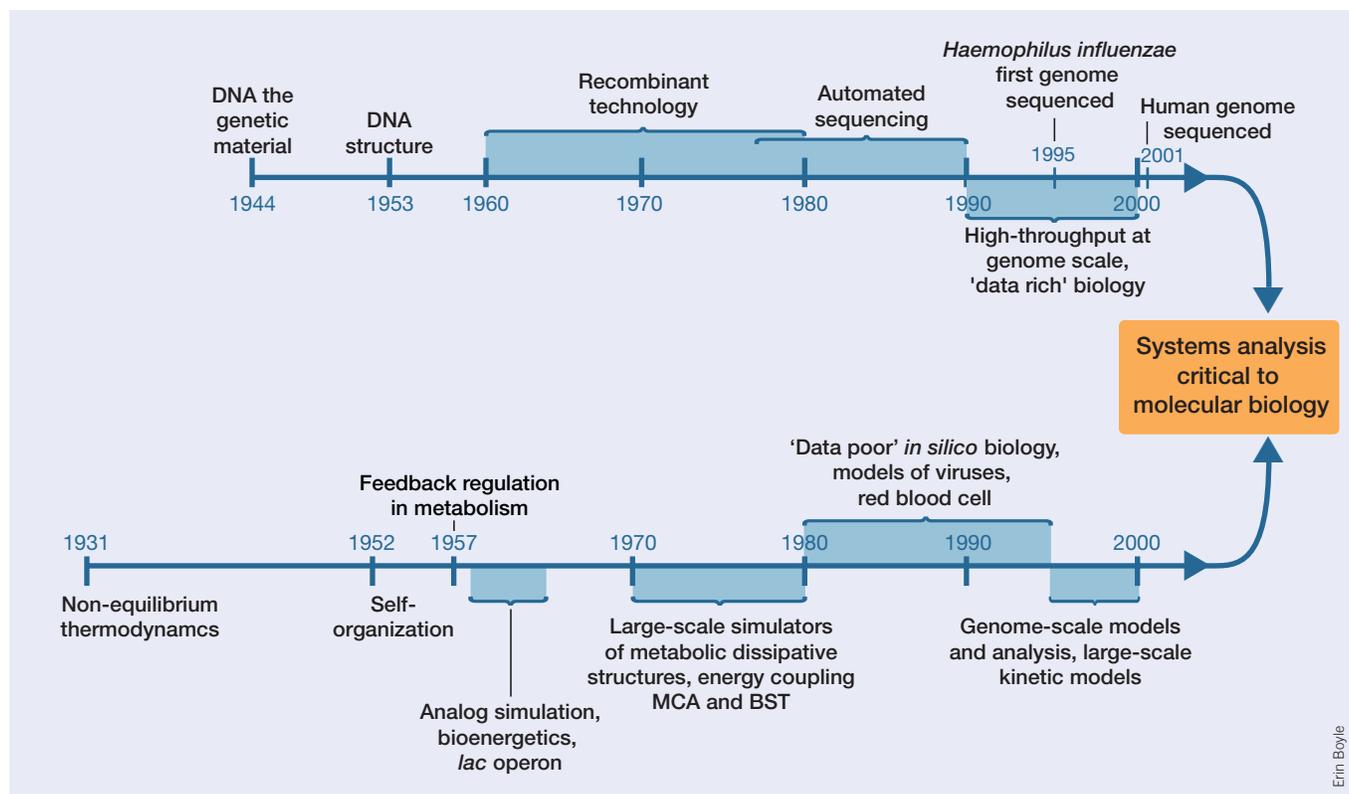
Figure 1 Two lines of inquiry led from the approximate onset of molecular biological thinking to present-day systems biology. The top timeline represents the root of systems biology in mainstream molecular biology, with its emphasis on individual macromolecules. Scaled-up versions of this effort then induced systems biology as a way to look at all those molecules simultaneously, and consider their interactions. The lower timeline represents the lesser-known effort that constantly focused on the formal analysis of new functional states that arise when multiple molecules interact simultaneously.

former driving force. Caplan, Essig and Rottenberg[12] later defined a coupling coefficient, which quantifies the extent to which two processes are coupled in a system and showed that this coefficient must range between 0 and 1.

These approaches were called nonequilibrium thermodynamics and constituted a prelude to systems biology at the cell and molecular levels in that they (i) dealt with integration quantitatively and (ii) aimed to discover general principles rather than just being descriptive. An improved procedure for describing ion movement and energy transduction in biological membranes, termed mosaic nonequilibrium thermodynamics, further progressed towards systems thinking in that it (iii) established a connection to molecular mechanisms and (iv) enabled the determination of the stoichiometry of membrane energy transduction from system data[13]. Peter Mitchell's[14] chemiosmotic coupling principle was another early case of systems analysis in cell and molecular biology. It stated that ATP synthesis was coupled in quite an indirect way to respiration, involving an entire intracellular system, including a volume surrounded by an ion-impermeable membrane and proton movement across it. Indeed, for eukaryotes, this provided much of the rationale for the organization of the mitochondrion. In his calculations verifying that that the proposed chemiosmotic mechanisms transferred sufficient free energy to empower ATP synthesis, Mitchell demonstrated the sort of quantitative thinking that would eventually prove crucial to the study of biochemical systems[14].

The problem of biological self-organization was to understand how structures, oscillations or waves arise in a steady and homogenous environment, a phenomenon called symmetry breaking. Turing[16] led the way, but the Prigogine school[17] and others developed the topic from the perspective of nonequilibrium thermodynamics in molecular contexts such as biochemical reactions involved in sugar metabolism (glycolysis). They demonstrated how having a sufficient number of nonlinearly interacting chemical processes in a single system such as the Zhabotinski reaction, a developing tissue, or glycolysis, could lead to symmetry-breaking as a result of self-amplification of random fluctuations. Of course, more recent molecular developmental biology studies have shown that reality is even more complicated; prespecification by external (maternally specified) gradients of morphogens may substitute for the random fluctuations, increasing the robustness of development[18]. Perhaps more importantly, Prigogine searched for and found a law (on minimum entropy production). Although it is strictly valid only in Onsager's near-equilibrium domain, it testified to the systems scientists' quest for the principles underlying systems, rather than just for their appearances.

Early on, oscillations in yeast glycolysis were the experimental systems of choice. Although intact cells were studied[19], more often measurements were made using cell extracts[20]. Reductionist biochemical thinking proclaimed that a single pacemaker enzyme should be responsible for the oscillations. Only relatively recently has systems-based analysis in one of our laboratories (H.V.W.) been used to reveal that the oscillations are simultaneously controlled by many steps in the intracellular network[21] and how the oscillations in the individual cells synchronize actively[22]. Of course, with the more recent experimental capability to inspect single cells dynamically, more and more cells are

seen to exhibit asynchronous oscillations of all sorts and some of these cases are up for systems biology analysis. Slime mold aggregation was another early case where a network of reactions was shown to be essential for systems biology reaching one step beyond cell biology, again by combining mathematical modeling with experimental molecular information[23].

Building large-scale models

Following the events of the late 1950s and early 1960s, researchers undertook efforts that were not well publicized and formulated mathematical models to simulate the functions of newly discovered regulatory circuits in cells. Even before digital computers became available, simulations of integrated molecular functions were performed on analog computers[24]. These efforts grew in scale to dynamic simulation of large metabolic networks in the 1970s[25–27]. Following the pathway-centered kinetic models in the seventies[28], cell-scale flux models of the human red cell were published by the late 1980s (ref. 29), and by the early 1990s genome-scale models of viruses and large-scale models of mitosis were formulated[30]. With the advent of genome-scale sequencing, the first genome-scale, constraint-based metabolic models for bacteria were constructed[31]. These models describe reconstructed networks and their possible functional states (phenotypes) and are now available at the genome-scale for a growing number of organisms. They treat the 'genome' as the 'system.'

Progress toward the development of detailed kinetic models at a large scale has proven to be slower. Some of these models approach computer replicas of pathways of metabolism, signal transduction and gene expression, and are active on the web, ready for experimentation and integration (compare http://www.siliconcell.net/). Obtaining *in vivo* numerical values for kinetic constants remains a key challenge.

Metabolic control analysis

We have agreed that contemporary systems biology has an historical root outside mainstream molecular biology, ranging from basic principles of self-organization in nonequilibrium thermodynamics, through large-scale flux and kinetic models to 'genetic circuit' thinking in molecular biology. 'Systems thinking' differs from 'component thinking' and requires the development of new conceptual frameworks.

Metabolic control analysis (MCA), developed in the early seventies[28,32], presented a key example of approaches to characterize properties of networks of interacting chemical reactions. At this time, thinking in biochemistry was dominated by the concept that there had to be a single 'rate-limiting' step at the beginning of all metabolic pathways. Criteria used to establish whether a given enzyme was rate-limiting referred to it as being far from equilibrium, strongly regulated by various metabolic factors or causing pathway flux to decrease when inhibited.

However, the application of these criteria to some metabolic pathways suggested that they contained more than a single rate-limiting step. Network thinking through MCA helped to resolve this paradox. First, mathematical models of metabolic pathways were developed both for inspiration and discovery, and subsequently used to check numerically the principles they conjectured[28,32]. Second, quantitative definitions were developed to describe the extent to which a step limited the flux through a pathway. This 'flux-control coefficient' of a particular step corresponded to the sensitivity coefficient of the pathway flux with respect to the activity of the particular enzyme. Third, these investigators looked for proof of the concept that there should be a single rate-limiting enzyme in a pathway that should have a flux-control coefficient of unity, with all others having flux control coefficients of

zero. Instead, they found a theorem stating that all the flux-control coefficients must sum to unity[28,32]. This result then suggested that there need not be a single rate-limiting step to a pathway and that instead many enzymes can contribute simultaneously to the control of the network. Thus, control was not a component property but a network property. The network nature of regulation was shown experimentally to be the case for mitochondrial ATP generation, where control was indeed distributed over more than three steps, and quite notably not particularly strong, neither for the first nor for the irreversible step of the pathway[33].

An important aspect of systems biology is to relate the system properties to the molecular properties of components that comprise a network. The kinetics-based sensitivity analysis by MCA, and its close relative, biochemical systems theory proposed by H.V.W and Chen[34], showed that by focusing on the properties of an individual component, one cannot properly decipher its role in the context of a whole network. The connectivity laws proven by MCA[28,34] (see other references in ref. 35) pinpointed how that distribution of control relates to network structure and the kinetic properties of all network components simultaneously. Similarly, the topological analyses of network structure by our groups[31,36] have revealed the existence of network-based definitions of pathways that can be used mathematically to represent all possible functional states of reconstructed networks[37]. Thus, a growing number of methods now exist to analyze the properties mathematically of the large-scale networks that we are now able to reconstruct based on high-throughput data.

Convergence

Figure 1 presents our interpretation of the history of systems analysis in cell and molecular biology. Events in the upper timeline have been much more to the fore of scientific thinking than those in the lower timeline. In one sense, the dazzling stream of discoveries and exciting technologies (most recently with genome-wide data) provides the 'biology' root to contemporary systems biology. In contrast, scientific progress in the lower timeline has never gained much notoriety, although work in this area was much more prominent in European science throughout this period. This latter branch might be thought of as the 'systems' root of systems biology.

Systems modeling and simulation in molecular biology was once seen as purely theoretical and not particularly relevant to understanding 'real' biology. However, now that molecular biology has become such a data-rich field, the need for theory, model building and simulation has emerged. The systems-directed root always had the ambition of discovering fundamental principles and laws, such as those of nonequilibrium thermodynamics and MCA. This ambition should now extend to systems biology.

All too often, the field has been perceived as just pattern recognition and phenomenological modeling. Systems biology is a thorough science with its own quest for scientific principles at the interface of physics, chemistry and biology, with its remarkable mixture of functionality, hysteresis, optimization and physical chemical limitations. *In silico* analysis of complex cellular processes (whether for data description, genetic engineering or scientific discovery), with its focus on elucidating system mechanisms, has in fact become critical for progress in biology.

The historical dichotomy in approaches to molecular biology must now be reconciled with the need to corral resources and expertise in systems approaches. Although the reductionist molecular biological root has been the focus of a plethora of investigations, literature sources and curricula, the same is not true for the systems molecular biology root. There is now a need for development of theoretical and

analytical approaches, curricula and educational materials to advance understanding of the systems in cell and molecular biology. Unknown to many, the 'pre-online PDF' era contains answers to many of the current challenges and pitfalls facing the field. So although systems biology has an intellectually exciting future ahead of it, the leaders in the field should try to minimize rediscovery and focus on the newer challenges facing us, particularly those that come with the application of existing concepts to genome-scale problems and identification of the new issues that arise from the study of cellular functions on this scale.

Where has this history brought us? We now have the growing and general recognition that systems analysis is important to the future evolution of cell and molecular biology. Some reeducation of workers in the field may be in order (http://www.systembiology.net/). Over the near term, it is likely that successes with practical applications of systems biology will be confined to unicellular systems. We are now seeing successful applications of systems biology to microbes, including pathway engineering (e.g., see our recent publications[37,38]), network-based drug design (e.g., H.V.W. and colleagues[39]), and prediction of the outcome of complex biological processes, such as adaptive evolution (B.O.P and colleagues[40]). Although the mathematical modeling of whole-body human systems cannot yet be linked to genome-wide data and models, data analysis and modeling are likely to contribute to the success of realizing the goal of individualized medicine. Even if we have to rely on less precise models than the currently available genome-scale models of microorganisms, systems biology may soon lead to better diagnosis and dynamic therapies of human disease than the qualitative methodology presently in use.

1. Umbarger, H.E. & Brown, B. Threonine deamination in *Escherichia coli*. II. Evidence for two L-threonine deaminases. *J. Bacteriol.* **73**, 105–12 (1957).
2. Yates, R.A. & Pardee, A.B. Control by uracil of formation of enzymes required for orotate synthesis. *J. Biol. Chem.* **227**, 677–692 (1957).
3. Beckwith, J.R. Regulation of the *lac* operon. Recent studies on the regulation of lactose metabolism in *Escherichia coli* support the operon model. *Science* **156**, 597–604 (1967).
4. Hunkapiller, T. *et al.* Large-scale and automated DNA sequence determination. *Science* **254**, 59–67 (1991).
5. Rowen, L., Magharias, G. & Hood, L. Sequencing the human genome. *Science* **278**, 605–607 (1997).
6. Scherf, M., Klingenhoff, A. & Werner, T. Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. *J. Mol. Biol.* **297**, 599–606 (2000).
7. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
8. Ge, H., Walhout, A.J. & Vidal, M. Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Genet.* **19**, 551–560 (2003).
9. Palsson, B.O. In silico biology through 'omics'. *Nat. Biotechnol.* **20**, 649–650 (2002).
10. Schrödinger, E. *What is life? The physical aspects of the living cell.* Based on Lectures Delivered under the Auspices of the Dublin Institute for Advanced Studies at Trinity College, Dublin, in February 1943. (Cambridge University Press, Cambridge, UK, 1944). http://home.att.net/~p.caimi/oremia.html
11. Onsager, L. Reciprocal relations in irreversible processes. *Phys. Rev.* **37**, 405–426 (1931).
12. Rottenberg, H., Caplan, S.R. & Essig, A. Stoichiometry and coupling: theories of oxidative phosphorylation. *Nature* **216**, 610–611 (1967).
13. Westerhoff, H.V. & Van Dam, K. Thermodynamics and Control of Biological Free-Energy Transduction (Elsevier, Amsterdam, 1987).
14. Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism. *Nature* **191**, 144–148 (1961).
15. Mitchell, P. Chemiosmotic Coupling in Oxidative and Photosynthetic Phosphorylation. (Glynn Research Ltd., Bodmin, UK, 1966).
16. Turing, A. The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. London, Ser. B* **237**, 37–72 (1952).
17. Glansdorff, P. & Prigogine, I. *Structure, Stabilité et Fluctuations* (Masson, Paris, 1971).
18. Lawrence, P.A. *The Making of a Fly* (Blackwell, London, 1992).
19. Chance, B., Estabrook, R.W. & Ghosh, A. Damped sinusoidal oscillations of cytoplasmic reduced pyridine nucleotide in yeast cells. *Proc. Natl. Acad. Sci. USA* **51**, 1244–1251 (1964).
20. Hess, B. & Boiteux, A. Oscillatory phenomena in biochemistry. *Annu. Rev. Biochem.* **40**, 237–258 (1971).
21. Teusink, B., Bakker, B.M. & Westerhoff, H.V. Control of frequency and amplitudes is shared by all enzymes in three models for yeast glycolytic oscillations. *Biochim. Biophys. Acta.* **1275**, 204–212 (1996).
22. Wolf, J. *et al.* Transduction of intracellular and intercellular dynamics in yeast glycolytic oscillations. *Biophys. J.* **78**, 1145–1153 (2000).
23. Tyson, J.J. & Murray, J.D. Cyclic AMP waves during aggregation of *Dictyostelium* amoebae. *Development* **106**, 421–426 (1989).
24. Goodwin, B.C. Oscillatory Organization in Cells, a Dynamic Theory of Cellular Control Processes (Academic Press, New York, 1963).
25. Garfinkel, D. *et al.* Computer applications to biochemical kinetics. *Annu. Rev. Biochem.* **39**, 473–498 (1970).
26. Loomis, W. & Thomas, S. Kinetic analysis of biochemical differentiation in *Dictyostelium discoideum*. *J. Biol. Chem.* **251**, 6252–6258 (1976).
27. Wright, B.E. The use of kinetic models to analyze differentiation. *Behavioral Sci.* **15**, 37–45 (1970).
28. Heinrich, R., Rapoport, S.M. & Rapoport, T.A. *Progr. Biophys. Mol. Biol.* **32**, 1–83 (1977).
29. Joshi, A. & Palsson, B.O. Metabolic dynamics in the human red cell. Part I—A comprehensive kinetic model. *J. Theor. Biol.* **141**, 515–528 (1989).
30. Novak, B. & Tyson, J.J. Quantitative analysis of a molecular model of mitotic control in fission yeast. *J. Theor. Biol.* **173**, 283–305 (1995).
31. Edwards, J.S. & Palsson, B.O. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.* **274**, 17410–17416 (1999).
32. Kacser, H. & Burns, J.A. *In Rate Control of Biological Processes* (ed., Davies, D.D.) 65–104 (Cambridge University Press, Cambridge, 1973).
33. Groen. A.K., Wanders, R.J.A., Van Roermund, C., Westerhoff, H.V. & Tager, J.M. Quantification of the contribution of various steps to the control of mitochondrial respiration. *J. Biol. Chem.* **257**, 2754–2757 (1982).
34. Savageau, M.A. *Biochemical Systems Analysis* (Addison-Wesley, Reading, MA, 1976).
35. Westerhoff, H.V. & Chen, Y. How do enzyme activities control metabolite concentrations? An additional theorem in the theory of metabolic control. *Eur. J. Biochem.* **142**, 425–430 (1984).
36. Westerhoff, H.V., Hofmeyr, J.H. & Kholodenko, B.N. Getting to the inside of cells using metabolic control analysis. *Biophys. Chem.* **50**, 273–283 (1994).
37. Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A. & Palsson, B.O. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **28**, 250–258 (2003).
38. Kholodenko, B.N. & Westerhoff, H.V. (eds.) *Metabolic Engineering in the Post Genomics Era* (Horizon Bioscience, UK, 2004).
39. Bakker, B.M. *et al.* Network-based selectivity of antiparasitic inhibitors. *Mol. Biol. Rep.* **29**, 1–5 (2002).
40. Ibarra, R.U., Edwards, J.S. & Palsson, B.O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420**, 186–189 (2002).