

Kunal Aggarwal
is a graduate student in the
School of Chemical and
Biomolecular Engineering at
Cornell University.

Kelvin H. Lee
is Associate Professor in the
School of Chemical and
Biomolecular Engineering at
Cornell University, and is the
Director of the Cornell
Proteomics Program.

Keywords: *systems biology,*
proteomics, microarrays

Functional genomics and proteomics as a foundation for systems biology

Kunal Aggarwal and Kelvin H. Lee

Date received (in revised form): 11th August, 2003

Abstract

Developments in high-throughput measurement technologies for biological molecules have created a paradigm shift in modern life science research. The field of systems biology attempts to provide a systems-level understanding by systematically organising the genomic, functional genomic and proteomic data obtained from genetic and environmental perturbations of interest and using the data to build a descriptive and mechanistic model of the biological phenomena. The goal is to build a mathematical framework with some predictive abilities. This review highlights the need for system-level understanding, lists some of the high-throughput measurement tools of importance in systems biology, reviews various types of experimental and computational approaches being used in systems biology research and attempts to address some of the challenges facing this research community.

INTRODUCTION

Life science-based research has undergone a revolutionary change in the past few years, with a shift in the focus of cellular studies with a reductionist approach towards an integrative approach. This shift has been driven by technology. The new integrative approach investigates 'complex' systems, which cannot be completely understood by investigation of individual components in isolation. Systems biology is a new field of science that develops a system-level understanding by describing quantitatively the interaction among all the individual components of the cell. The ultimate aim of such an approach is to develop computational models of these complex systems so that the response of the biological system to any kind of perturbation, for example, environmental disturbance, genetic mutation etc, can be predicted.^{1,2}

The reductionist approach to biological research, which has been extremely important to the development of a basic understanding of living system, is geared towards identifying the individual components (genes, proteins, metabolites

etc) responsible for a particular phenomenon in an organism (eg metabolic activity, response to external stimuli etc). This approach has proven effective at elucidating key molecular components of living systems, leading to a variety of important applications in agriculture and medicine. It is now clear, however, that information at only one level (the genome or the proteome, for example) by itself cannot fully explain the behaviour of any particular biological system. An often cited observation is the lack of obvious correlation between protein and mRNA abundance levels in yeast³ and human liver cells.⁴ The changes induced in protein expression do not correspond linearly to the changes in mRNA expression.⁵ Furthermore, it has been shown mathematically that a detailed understanding of the control of even the simplest gene networks requires information at both the mRNA and protein expression levels.⁶ Biological systems also show properties, such as robustness, redundancy, modularity and regulation via feedback control, characteristic of complex systems.⁷ Cells

Kelvin H. Lee,
Cornell University,
School of Chemical and
Biomolecular Engineering,
120 Olin Hall,
Ithaca, NY 14853-520,
USA

Tel: +1 607 255 4215
Fax: +1 607 255 9166
E-mail: KHL9@cornell.edu

Systems biology involves an iterative process of experimental observations and mathematical modelling

can be insensitive to many mutations, implying that the failure of many genes or specific regulatory interactions does not have a dominant effect on an observed phenotype unless a certain set of genes is simultaneously altered.⁸ Thus, there is a clear need for an integrated approach based on the use of simultaneous measurements of functional genomic, proteomic and metabolomic parameters to study and understand the biological systems.

The concept of an integrative study of biological systems is not particularly new. It was probably first proposed by Norbert Wiener in 1948, while defining the area of cybernetics.⁹ Subsequently, a number of investigators took a systems-level approach to define a set of principles that linked the behaviour of living matter to system characteristics and functions.^{10,11} A lack of understanding at the molecular level, however, hindered such attempts, as one could only describe the biological systems at the physiological level. With the emergence of high-throughput technologies for molecular level measurements of gene expression, protein expression and protein–protein interactions, and with the availability of complete genome sequences of various species, it is now possible to measure a large number of cellular components simultaneously and to perform system-level studies at the molecular level.

Systems biology is still an emerging field, and various attempts have been made to identify the key areas that need to be investigated to achieve a system-level understanding of a biological system. While there is no definition of a systems biology approach, one possible paradigm for such studies is presented in Figure 1. In particular, it is important to define all of the components of the system, including the regulatory relationships between genes and interactions of proteins and biochemical pathways, and to use this knowledge to formulate a primitive model (biochemical or mathematical). Once the system structure is defined, system behaviour can be

analysed further using specific genetic and/or environmental perturbations. The data generated from such an analysis can either be integrated with the initial model or used to refine the model, such that its predictions are consistent with the experimental observations. Multiple hypotheses or several sets of parameters for a particular model may help to predict the observed system behaviour and one may require new perturbations and measurements eliciting different system responses between models to discriminate between them. Thus, systems biology involves an iterative process of matching experimental observations against model predictions to formulate new models and new experiments to test them.^{1,12} Other important areas of systems biology research include identification of mechanisms that control the state of the cells so that they can be used to identify potential therapeutic targets for treatment of diseases. Eventually, these detailed studies should enable researchers to construct biological systems having the desired properties based on definite design principles and simulations.¹³

This review will emphasise how the data collected from functional genomics and proteomics experiments may be used to develop computational frameworks that organise and describe biological phenomena.

TOOLS AND RESOURCES

Systematic characterisation of cellular responses requires quantitative biological measurements that can yield information about cellular processes at multiple levels. This section lists some of the experimental tools which can be, and are currently being, used to generate systems-level data. For a detailed description of the techniques mentioned, the reader is directed to the appropriate references, as such descriptions are beyond the scope of this article. Only a brief description will be provided here, to familiarise the reader with some of the tools that have become important in systems biology.

The expression levels of thousands of

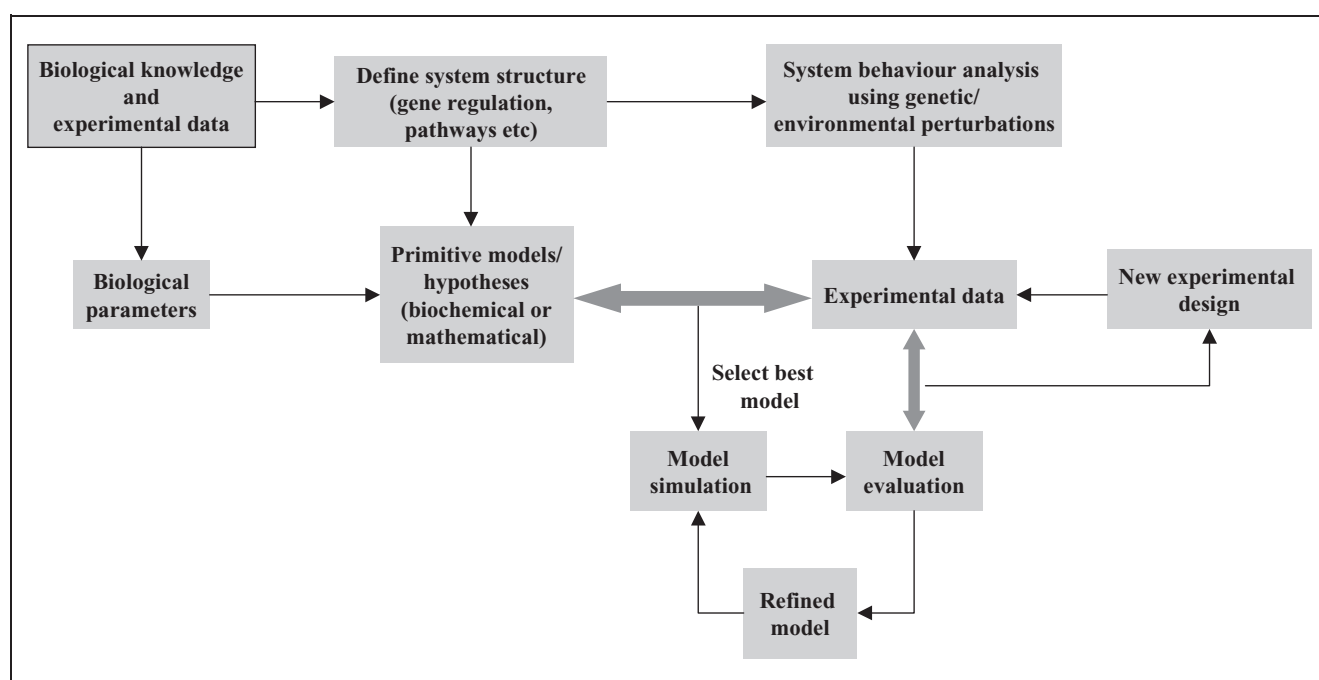


Figure 1: In one view of systems biology, the previous biological knowledge and experimental data are used to define the relationships between various components of the system of interest and to extract biological parameters. These relationships and parameters can be used to build primitive biochemical or mathematical models to describe a particular biological phenomenon. A comparison between the experimental data generated from perturbations of the system and predictions from various candidate models can be used to screen for the model or a set of parameters that best describes the data. One may also design new experiments eliciting different system responses between models to discriminate among candidate models. The experimental data thus generated can be used either to refine the model by integrating the data into the model or to evaluate the model simulation results. This iterative process of matching experimental observations against model predictions can continue as appropriate

DNA microarrays, 2D gels and mass spectrometry are important tools in systems biology

genes can be measured and analysed simultaneously using DNA microarrays (spotted arrays¹⁴ and high density arrays¹⁵). Transcribed mRNA can also be analysed using alternative methods such as serial analysis of gene expression,¹⁶ differential display¹⁷ and cDNA fingerprinting.¹⁸ DNA microarrays are more popular, however, because they can be used to resequence whole stretches of DNA, identify single nucleotide polymorphisms and characterise spliced genes, in addition to their important role in mRNA expression analysis.

The proteins within a cell are commonly analysed using two-dimensional (2D) gel electrophoresis, in which complex mixtures are resolved first by isoelectric point and then by size on a polyacrylamide gel.¹⁹ This technique can

also be used to detect certain post-translational modifications. This approach, however, like most protein expression technologies, suffers from an inability to quantify all proteins effectively and from problems resolving certain classes of proteins.²⁰ The proteins that appear as individual spots on a 2D gel can be further analysed using mass spectrometry (MS), such as matrix assisted laser desorption/ionisation time of flight (MALDI-TOF) MS, in which the masses of resulting peptides from a proteolytic digest are compared with *in silico* digests of protein databases to characterise the protein spot.²¹ Complex protein samples can also be analysed using liquid chromatography-tandem mass spectrometry (LC-MS/MS), in which protein and peptide mixtures are supplied

to the mass spectrometer from a high-performance liquid chromatography (HPLC) system.^{20–22}

The relative quantification of protein expression can be measured using isotopic dilution strategies on a mass spectrometer (for example, isotope coded affinity tags, or ICATs).²³ Protein–protein and protein–small molecule interactions can be studied using yeast two hybrid screening,²⁴ co-immunoprecipitation,²⁵ surface plasmon resonance²⁶ and fluorescence resonance energy transfer (FRET).²⁷ Spotted protein arrays can also be used to probe protein–protein and protein–DNA interactions as well as protein activity.^{28,29} The cellular metabolite concentrations can be quantified using MS.³⁰ All of these tools can be used in various permutations and combinations to obtain different types and levels of information about the biological system. Irrespective of the method of data collection, an ideal study involves the simultaneous measurement of all the genes and proteins related to those of interest from the biological samples. To be effective as a systems biology approach, data should also be collected at several time points to capture time-dependent dynamic phenomena. The quality of the data produced needs to be controlled if it is to be used as a reference point for simulations and modelling.

Sequencing efforts, as well as genomic and proteomic technologies, have generated vast amounts of biological data that has been systematically organised into several online databases. The National Center for Biotechnology Information,³¹ Protein Data Bank,³² Kyoto Encyclopedia of Genes and Genomes,³³ Biomolecular Interaction Network Database³⁴ and Encyclopedia of *Escherichia coli* Genes and Metabolism³⁵ are some of the many databases that contain information on nucleic acid and amino acid sequences, molecular properties, interactions and pathways. These databases provide an essential infrastructure for the ongoing and future efforts on modelling of biological systems. They cannot achieve

maximum impact, however, unless a standard format for storage of information is available, so that data can be searched easily and transported across various software platforms. The Systems Biology Markup Language (SBML)³⁶ and CellML³⁷ are examples of such standards that enable exchange of data and modelling information among a wide variety of software systems. Beyond the development of databases and standards, efforts are also underway to create a comprehensive and highly integrated simulation analysis environment for systems biology research. The Systems Biology Workbench project aims to create an integrated, easy to use software environment that enables sharing of models and resources between simulation and analysis tools for systems biology.³⁸ E-cell provides modelling and simulation environment for biochemical and genetic processes.³⁹

SYSTEMS BIOLOGY APPROACHES

As mentioned above, systems biology is a new field; significant research efforts are focused towards identifying the structure of biological systems. For example, microarray technology has been used to conduct a thorough analysis of gene expression patterns in *E. coli* cultured in different environmental conditions.⁴⁰ Although *E. coli* physiology has been extensively studied for decades, genome-wide expression profiling has helped to discover new features regarding the response of *E. coli* to poor carbon source availability. In one such study, cells cultured in acetate showed induction of the malic enzyme-*ppsA* and the glycolate pathway and repression of glycolytic and glucose phosphotransferase genes when compared with those grown in glucose. The same work also provides experimental evidence for up-regulation of metabolic genes and down-regulation of genes involved in cell replication, transcription and translation machinery in *E. coli* when cultured in a poor carbon source.⁴¹

Standard formats for information storage are necessary for sharing and searching data

Systems-level approaches are often used to study gene expression regulation. These studies may involve the perturbation of each known essential component in a defined biochemical pathway using a genetic mutation/deletion strategy and/or by a change in the environment. One example of this approach is the study of galactose metabolism in yeast, where the understanding of the existing pathway for galactose utilisation (GAL) was refined on the basis of the above-mentioned perturbation experiment results.⁴² Wild-type and genetically altered yeast strains with a complete deletion of one of the GAL genes involved in transport, enzymatic or regulatory function were grown in the presence and absence of galactose, and global changes in mRNA expression resulting from each perturbation were examined. Expression data from each perturbation were then incorporated into a network model of known molecular interactions connecting galactose utilisation with other metabolic processes in yeast, generated using publicly available databases, to increase its predictive power. The model was further refined using data from additional perturbations, like double gene deletions. Studies of a similar nature have been conducted in *Halobacterium sp.* to reveal a coordinated co-regulation of several interconnected biochemical pathways for phototrophy.⁴³ These types of studies not only help in understanding the biological systems better, by mapping the interactions between various genes and gene products, but also aid in identifying potential drug targets. A recent gene and protein expression profiling of *Mycobacterium tuberculosis* under conditions of nutrient starvation has revealed induction of some genes, which may facilitate survival of *M. tuberculosis* under stringent conditions and which may represent relevant drug targets.⁴⁴ Although a diagrammatic representation of the information on a pathway facilitates the understanding of the network topology and identification of drug

targets, its capacity to predict cell behaviour in response to an environmental or genetic change is very limited. Mathematical models of biological systems should summarise the current information into a coherent whole. They help in understanding some of the qualitative features of the biological systems and can help in refining the existing hypothesis about a biological process.⁴⁵

Computational biology has already proven its usefulness in the field of biochemical engineering. Cybernetic models, which hypothesise that cells always function and allocate resources to optimise the resulting outcome, have been successfully used to quantitatively describe complex biological phenomena such as the dynamics of mixed substrate utilisation⁴⁶ and the functioning of storage pathways.⁴⁷ An *E. coli* single cell model dynamically couples information from various fields, like biochemistry, microbiology, physiology, genetics etc, to simulate the growth of a single bacterial cell, and has served as a platform for further development to address other questions.⁴⁸ Thus, the application of mathematical structures and methods in the field of systems biology is essential if it has to be able to predict and design the dynamic behaviour of complex biological systems.

Two of the key issues which need to be addressed for this field to continue to move forward are access to appropriate data from functional genomics and proteomic experiments and the development of useful modelling frameworks to incorporate this information. Accordingly, there is an increasing dependence on the use of mathematical structures to organise data and model biological phenomena. Computational methods have been used to perform a systematic analysis of interactions between signalling pathways. In one such study,⁴⁹ the interactions between the cyclic adenosine monophosphate (cAMP) signalling pathway, Ca²⁺/calmodulin-dependent

Mathematical models help in describing and predicting dynamic behaviour of biological systems

protein kinase II (CamKII) regulation and epidermal growth factor receptor (EGFR) signalling pathways resulted in the formation of complex networks with non-intuitive emergent properties that are not seen in individual pathways, for example, integration of signals across multiple time scales, generation of distinct outputs depending on input strength and duration, and self-sustaining feedback loops which may result in bi-stable behaviour with discrete steady-state activities. Considerable insights have been gained in the EGFR system trafficking dynamics by combining quantitative experimental and modelling approaches. Autocrine signalling is an excellent example of the application of a mathematical model to a seemingly complex system to show that it operates by simple principles.⁵⁰

Modelling approaches in systems biology

The modelling approaches used in systems biology can be broadly classified as qualitative and mechanistic.⁵¹ Qualitative, or hypothesis-driven, models are usually logical or statistical models of selected biological phenomena that fit known information. They have mostly been used to explore gene networks using the assignment of Boolean parameters to individual connections; however, complex networks are now being simulated in a more realistic fashion using fuzzy logic, an extension of Boolean logic.⁵² Statistical models like neural networks and Bayesian models can be used to simulate complex biological mechanisms whenever perturbation response or time series data are available. Mechanistic, or data-driven, models are often based on data from high-throughput experiments. Clustering methods have been used with DNA microarray data to identify groups of genes that may be related.⁵³ Although such methods can handle large-scale profile data, they cannot directly deduce network structures and cannot infer possible modifications and translational control.¹³ Experimental data from various sources, such as expression profiles, protein–protein interactions and others, needs to be

integrated to obtain a greater understanding. ‘Transcriptome–interactome correlation mapping’⁵⁴ has been used to compare the interactions between proteins encoded by genes that belong to common expression profiling clusters in *Saccharomyces cerevisiae* with those between proteins encoded by genes that belong to different clusters. Such integration of data has revealed that genes with similar expression profiles have a greater likelihood of encoding interacting proteins.

Once formulated, these models can be analysed for their properties using either simulation or standard analysis approaches, such as flux balance analysis (FBA), metabolic control analysis, stoichiometric network analysis, sensitivity analysis and bifurcation theory.⁵¹ The capabilities of a reconstructed metabolic network can be analysed using FBA, and the inter-relationships can be understood using sensitivity analysis. Bifurcation analysis can be used to understand the dynamics of the system, potential operating points and phase transition boundaries.⁵¹ Elementary mode analysis links structural analysis to metabolic flux analysis and helps in investigating all the physiological states that are meaningful for the cell in the long-term perspective.⁵⁵

In silico modelling approaches are also used in the analysis of biological processes. These use known biological relationships and real biological data derived from biochemical properties of gene products to create virtual systems by incorporating the data in a numerical format into a set of equations describing the system to be simulated. The underlying assumptions in a model can be assessed using simulations to predict the system dynamics and testing these predictions against *in vitro* and *in vivo* studies.⁵⁶ Systems have been generated to simulate the development of heart⁵⁷ and several organs in *Drosophila*.⁵⁸

Even with so many tools in hand, computational biology has not yet evolved to a stage where it can address some of the problems of particular

Effective modelling efforts do not require detailed knowledge of all system parameters

relevance (eg curing cancer). One significant reason for this is the belief that all of the quantitative parameters for all of the components and interactions in the cell need to be known beforehand to formulate a mathematical framework for a biological process. Many results available for certain biological systems, and the mathematical frameworks describing them, however, require relatively little or no information on quantitative values of system parameters.⁵⁹ Flux balance analysis has been used to deal with the lack of kinetic information and to impose known systemic constraints on the function of entire metabolic network in *E. coli* MG1655. The set of allowable metabolic flux distributions within the imposed systemic constraints has been probed for the optimal flux distributions using linear programming.⁶⁰ A cybernetic regulatory program managing gene expression, translation and enzyme activity has been successfully used as a surrogate for missing molecular details to simulate the management of *E. coli* central carbon metabolism.⁶¹ Statistical mechanics has been used to extract predictions from the eukaryotic signal transduction model, even in the face of indeterminacy of parameters and network topology (Kevin Brown, personal communication). Investigators are also dedicated to the identification of the minimum and the most important information required to describe biological systems. The bacterial genomes of *Mycoplasma genitalium* and *Haemophilus influenzae* have been compared to define the minimal gene set that is necessary and sufficient for supporting cellular life.⁶² A computational procedure has been used to identify the minimal set of metabolic reactions in *E. coli* capable of supporting given targets on the growth rate for different substrates by simultaneously considering the effect of all reactions on cell growth.⁶³ These efforts can help in building the model of a primitive cell that may have existed in the early stages of the evolution of life. A lack of complete data sets and knowledge about network topology is an important

issue that systems biology will need to address to continue moving forward.

CLOSING REMARKS

Systems biology is an emerging field, poised to contribute substantially to our overall understanding of biological phenomena, including all of the important applications in agriculture, technology and medicine that have been pursued or considered. The combination of data from various '-omic' technologies with improved data management architectures to build reliable computational frameworks will cause a shift in emphasis from wet lab techniques to include *in silico* experiments. This will make efforts more efficient, leading to shortened research and development time and reducing the associated costs. The results from such computations can then be used to refine and design a further stage of experiments. Such changes may significantly reduce the investment required to develop new drugs because of increasing reliance on *in silico* studies. Thus, systems biology may become the dominant perspective in future life science research in academic and applied sciences. There is a critical need, however, for scientists, students and investigators capable of interacting beyond traditional disciplines, because many of the important contributions may be based on multidisciplinary efforts. Systems biology therefore not only requires integration of information about organisms and cells at multiple levels but also necessitates the integration of knowledge from biological science and tools from mathematics and computer science for a better understanding of the biological systems.

Acknowledgments

This work is supported by National Science Foundation BES-9874938 and BES-0120315.

References

1. Kitano, H. (2002), 'Systems biology: a brief overview', *Science*, Vol. 295, pp. 1662–1664.

2. Wolkenhauer, O. (2001), 'Systems biology: The reincarnation of systems theory applied in biology?', *Brief. Bioinformatics*, Vol. 2, pp. 258–270.
3. Gygi, S. P., Rochon, Y., Franza, B. R. and Aebersold, R. (1999), 'Correlation between protein and mRNA abundance in yeast', *Mol. Cell. Biol.*, Vol. 19, pp. 1720–1730.
4. Anderson, L. and Seilhamer, J. (1997), 'A comparison of selected mRNA and protein abundances in human liver', *Electrophoresis*, Vol. 18, pp. 533–537.
5. Griffin, T. J., Gygi, S. P., Ideker, T. *et al.* (2002), 'Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*', *Mol. Cell. Proteomics*, Vol. 1, pp. 323–333.
6. Hatzimanikatis, V. and Lee, K. H. (1999), 'Dynamical analysis of gene networks requires both mRNA and protein expression information', *Metab. Eng.*, Vol. 1, pp. E1–E7.
7. Kitano, H. (2002), 'Looking beyond the details: A rise in system-oriented approaches in genetics and molecular biology', *Curr. Genet.*, Vol. 41, pp. 1–10.
8. Bailey, J. E. (1999), 'Lessons from metabolic engineering for functional genomics and drug discovery', *Nat. Biotechnol.*, Vol. 17, pp. 616–618.
9. Wiener, N. (1948), '*Cybernetics or Control and Communication in the Animal and the Machine*', The MIT Press, Cambridge, MA.
10. Bertalanffy, L. V. (1969), '*General System Theory, Foundations, Development, Applications*', George Braziller, New York, NY.
11. Mesarovic, M. D. (1968), '*System Theory and Biology*', Springer-Verlag, New York, NY.
12. Ideker, T., Galitski, T. and Hood, L. (2001), 'A new approach to decoding life: Systems biology', *Annu. Rev. Genomics Hum. Genet.*, Vol. 2, pp. 343–372.
13. Kitano, H. (2000), 'Perspectives on systems biology', *New Generation Computing*, Vol. 18, pp. 199–216.
14. Lockhart, D. J. and Winzler, E. A. (2000), 'Genomics, gene expression and DNA arrays', *Nature*, Vol. 405, pp. 827–836.
15. Selinger, D. W., Cheung, K. J., Mei, R. *et al.* (2000), 'RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array', *Nature Biotechnol.*, Vol. 18, pp. 1262–1268.
16. Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. (1995), 'Serial analysis of gene-expression', *Science*, Vol. 270, pp. 484–487.
17. Liang, P. and Pardee, A. B. (1992), 'Differential display of eukaryotic messenger-RNA by means of the polymerase chain-reaction', *Science*, Vol. 257, pp. 967–971.
18. Ivanova, N. B. and Belyavsky, A. V. (1995), 'Identification of differentially expressed genes by restriction endonuclease-based gene-expression fingerprinting', *Nucleic Acids Res.*, Vol. 23, pp. 2954–2958.
19. O'Farrell, P. H. (1975), 'High-resolution 2-dimensional electrophoresis of proteins', *J. Biol. Chem.*, Vol. 250, pp. 4007–4021.
20. Lee, K. H. (2001), 'Proteomics: a technology-driven and technology-limited discovery science', *Trends Biotechnol.*, Vol. 19, pp. 217–222.
21. Aebersold, R. and Mann, M. (2003), 'Mass spectrometry-based proteomics', *Nature*, Vol. 422, pp. 198–207.
22. Patterson, S. D. and Aebersold, R. H. (2003), 'Proteomics: The first decade and beyond', *Nat. Genet.*, Vol. 33, pp. 311–323.
23. Gygi, S. P., Rist, B., Gerber, S. A. *et al.* (1999), 'Quantitative analysis of complex protein mixtures using isotope-coded affinity tags', *Nat. Biotechnol.*, Vol. 17, pp. 994–999.
24. Bartel, P. L., Roecklein, J. A., SenGupta, D. and Fields, S. (1996), 'A protein linkage map of *Escherichia coli* bacteriophage T7', *Nat. Genet.*, Vol. 12, pp. 72–77.
25. Herrmann, J. M., Westermann, B. and Neupert, W. (2001), 'Analysis of protein-protein interactions in mitochondria by coimmunoprecipitation and chemical cross-linking', *Methods Cell Biol.*, Vol. 65, pp. 217–230.
26. Smith, E. A., Thomas, W. D., Kiessling, L. L. and Corn, R. M. (2003), 'Surface plasmon resonance imaging studies of protein-carbohydrate interactions', *J. Am. Chem. Soc.*, Vol. 125, pp. 6140–6148.
27. Kenworthy, A. K. (2001), 'Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy', *Methods*, Vol. 24, pp. 289–296.
28. Lueking, A., Horn, M., Eickhoff, H. *et al.* (1999), 'Protein microarrays for gene expression and antibody screening', *Anal. Biochem.*, Vol. 270, pp. 103–111.
29. Walter, G., Bussow, K., Cahill, D. *et al.* (2000), 'Protein arrays for gene expression and molecular interaction screening', *Curr. Opin. Microbiol.*, Vol. 3, pp. 298–302.
30. Fiehn, O. (2002), 'Metabolomics — the link between genotypes and phenotypes', *Plant Mol. Biol.*, Vol. 48, pp. 155–171.

31. URL: <http://www.ncbi.nlm.nih.gov> (last accessed on 06/30/03).
32. URL: <http://www.rcsb.org/pdb> (last accessed on 06/30/03).
33. URL: <http://www.genome.ad.jp/kegg> (last accessed on 06/30/03).
34. URL: <http://www.isc.org/products/BIND> (last accessed on 06/30/03).
35. URL: <http://ecocyc.org/> (last accessed on 06/30/03).
36. Hucka, M., Finney, A., Sauro, H. M. *et al.* (2003), 'The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models', *Bioinformatics*, Vol. 19, pp. 524–531.
37. URL: <http://www.cellml.org> (last accessed on 06/30/03).
38. Hucka, M., Finney, A., Sauro, H. *et al.* (2001), 'The ERATO systems biology workbench: an integrated environment for multiscale and multitheoretic simulations in systems biology', in Kitano, H., Ed., '*Foundations of Systems Biology*', The MIT Press, Cambridge, MA, pp. 125–143.
39. Tomita, M., Hashimoto, K., Takahashi, K. *et al.* (1999), 'E-CELL: software environment for whole-cell simulation', *Bioinformatics*, Vol. 15, pp. 72–84.
40. Oh, M. K. and Liao, J. C. (2000), 'Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*', *Biotechnol. Prog.*, Vol. 16, pp. 278–286.
41. Oh, M. K., Rohlin, L., Kao, K. C. and Liao, J. C. (2002), 'Global expression profiling of acetate-grown *Escherichia coli*', *J. Biol. Chem.*, Vol. 277, pp. 13175–13183.
42. Ideker, T., Thorsson, V., Ranish, J. A. *et al.* (2001), 'Integrated genomic and proteomic analyses of a systematically perturbed metabolic network', *Science*, Vol. 292, pp. 929–934.
43. Baliga, N. S., Pan, M., Goo, Y. A. *et al.* (2002), 'Coordinate regulation of energy transduction modules in *Halobacterium sp.* analyzed by a global systems approach', *Proc. Natl. Acad. Sci. USA*, Vol. 99, pp. 14913–14918.
44. Betts, J. C., Lukey, P. T., Robb, L. C. *et al.* (2002), 'Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling', *Mol. Microbiol.*, Vol. 43, pp. 717–731.
45. Bailey, J. E. (1998), 'Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities', *Biotechnol. Prog.*, Vol. 14, pp. 8–20.
46. Ramakrishna, R., Ramakrishna, D. and Konopka, A. E. (1996), 'Cybernetic modeling of growth in mixed, substitutable substrate environments: Preferential and simultaneous utilization', *Biotechnol. Bioeng.*, Vol. 52, pp. 141–151.
47. Varner, J. and Ramakrishna, D. (1998), 'Application of cybernetic models to metabolic engineering: Investigation of storage pathways', *Biotechnol. Bioeng.*, Vol. 58, pp. 282–291.
48. Domach, M. M. and Shuler, M. L. (1984), 'A finite representation model for an asynchronous culture of *E. coli*', *Biotechnol. Bioeng.*, Vol. 26, pp. 877–884.
49. Bhalla, U. S. and Iyengar, R. (1999), 'Emergent properties of networks of biological signaling pathways', *Science*, Vol. 283, pp. 381–387.
50. Wiley, H. S., Shvartsman, S. Y. and Lauffenburger, D. A. (2003), 'Computational modeling of the EGF-receptor system: A paradigm for systems biology', *Trends Cell Biol.*, Vol. 13, pp. 43–50.
51. Arkin, A. P. (2001), 'Synthetic cell biology', *Curr. Opin. Biotechnol.*, Vol. 12, pp. 638–644.
52. Thieffry, D. and Thomas, R. (1998), 'Qualitative analysis of gene networks', *Proc. Pac. Symp. Biocomput.*, Vol. 3, pp. 77–88.
53. Michaels, G. S., Carr, D. B., Askenazi, M. *et al.* (1998), 'Cluster analysis and data visualization of large-scale gene expression data', *Proc. Pac. Symp. Biocomput.*, Vol. 3, pp. 42–53.
54. Ge, H., Liu, Z. H., Church, G. M. and Vidal, M. (2001), 'Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*', *Nat. Genet.*, Vol. 29, pp. 482–486.
55. Stelling, J., Klamt, S., Bettenbrock, K. *et al.* (2002), 'Metabolic network structure determines key aspects of functionality and regulation', *Nature*, Vol. 420, pp. 190–193.
56. Kitano, H. (2002), 'Computational systems biology', *Nature*, Vol. 420, pp. 206–210.
57. Noble, D. (2002), 'Modeling the heart — from genes to cells to the whole organ', *Science*, Vol. 295, pp. 1678–1682.
58. Kyoda, K. and Kitano, H. (1999), 'Simulation of genetic interaction for *Drosophila leg* formation', *Proc. Pac. Symp. Biocomput.*, Vol. 4, pp. 77–89.
59. Bailey, J. E. (2001), 'Complex biology with no parameters', *Nat. Biotechnol.*, Vol. 19, pp. 503–504.
60. Edwards, J. S., Ibarra, R. U. and Palsson, B. O. (2001), '*In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with

- experimental data', *Nat. Biotechnol.*, Vol. 19, pp. 125–130.
61. Varner, J. D. (2000), 'Large-scale prediction of phenotype: Concept', *Biotechnol. Bioeng.*, Vol. 69, pp. 664–678.
62. Mushegian, A. R. and Koonin, E. V. (1996), 'A minimal gene set for cellular life derived by comparison of complete bacterial genomes', *Proc. Natl. Acad. Sci. USA*, Vol. 93, pp. 10268–10273.
63. Burgard, A. P., Vaidyaraman, S. and Maranas, C. D. (2001), 'Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments', *Biotechnol. Prog.*, Vol. 17, pp. 791–797.