Preface

This issue of the *IBM Journal of Research and Development* focuses on the growing field of systems biology. While the precise definition of the term "systems biology" is continually undergoing refinement and is open for discussion, researchers agree that the vast quantities of data generated by high-throughput biology, and related techniques used to study living systems, ensure that computation and computers will play an ever-increasing role in biological research. This computer-based approach gives systems biology researchers the opportunity to investigate the ways in which various kinds of data interrelate in order to go beyond an understanding of individual parts to an understanding of the function of larger biological systems. Using computational tools, the computational biologist analyzes, models, and simulates biological systems. Biology, once considered primarily as an observational science, has become quite quantitative.

While computation itself provides a crucial tool for understanding biological systems, an important interplay also exists between computer modeling and simulation and "wet" laboratory experimentation. Each kind of investigation focuses researchers' attention and can be used to validate the findings from the other. The iterative and comparative process between laboratory experimentation and computer experimentation can increase research productivity as the computation guides the experiment and the experiment is used to validate and refine the computation.

Given this background, in this *Journal* issue we view systems biology from a broad perspective and have assembled a collection of papers that reflect computational issues associated with several areas of systems biology. Many of the papers highlight computational tools that researchers have developed to support systems biology efforts. This emphasis on systems biology is changing the way that researchers look at the biological world today. Researchers now create the algorithms and computational ecosystems that help investigate and explore biological systems. They are helping biological and chemical experimentalists gain new insights and explore different scenarios in biology. The remarkable panoply of computational approaches and perspectives forms what we believe to be the essence of systems biology.

The introductory paper, by Burbeck and Jordan, explores systems biology and speculates on its impact, using information gleaned from numerous interviews with expert computational biologists. The paper gives an overview of the key ideas underlying systems biology and helps place the remaining papers in context.

As we have already emphasized, high-throughput biology is generating vast quantities of data, and much of this data resides in databases. The task of gaining insight from the data as it relates to complex biological systems can be daunting. To make this task easier, researchers are developing a variety of tools to represent and study such data. The paper by Eckman and Brown describes a tool used to handle queries relating to biological networks. This tool, the Systems Biology Graph Extender, extends the IBM DB2® database so that it can deal effectively with queries on biological graph objects. On a related topic, the Edinburgh Pathway Editor is a tool that combines visualization, editing, and database manipulation of information related to biological networks. This open-source software is described and compared with other pathway editors in the paper by Sorokin et al.

The paper by Podowski et al. also approaches the interpretation of biological data through appropriate visualization tools. Their paper discusses the use of *heatmaps* or pattern-revealing aggregate views of the data that allow information to be displayed in representations that may involve more than two dimensions. In particular, a parallel-dataset heatmap viewer permits four-dimensional data to be displayed. Such tools are helping biologists understand and gain insight from large quantities of genomic data.

Computational modeling of biological systems can be especially challenging because biological phenomena occur at vastly different time and length scales. For example, chemical reactions within a cell may result in the contraction of muscle tissue at a much larger size scale. This multi-scale and multi-physics nature of biology often requires substantial computational power to resolve the finer scales of the models. The paper by Hussan et al. presents a computational model that bridges spatial scales to develop our understanding of the way protein complexes affect cardiac muscle responses. Similarly, the papers by Peirce et al. and Nickerson et al. describe multiple scale issues and illustrate ways in which researchers are making strides to incorporate multi-scale and multi-physics perspectives and parameters into complex models, such as models of the human heart.

A fundamental issue associated with systems biology, and especially with gene network analysis, is the understanding of transcription regulation. In particular, gene expression is modulated by protein transcription factors. A first step in understanding transcriptional regulation requires the mapping of transcription factors to the genes they regulate and to the particular nucleotide sequences to which they bind. This area of study is a particularly apt example of systems biologists working with a "parts list" and determining the overall function of the parts when they are working together. Holloway et al. describe the use of support vector machine algorithms, with a penalty for misclassification, to investigate

classification schemes for the prediction of transcription factor binding, and the researchers demonstrate the robustness of their method.

As the systems biology discipline evolves, researchers who are newcomers from other physical science disciplines investigate approaches to understanding biology that have been successful in these other disciplines. The paper by Schoeberl et al. focuses on Model-Based Design (MBD), which has proven successful in the automotive, chemical, and aerospace industries. These authors indicate that with improvement of computations for modeling biological systems, MBD should provide an attractive approach for designing new drugs and for the *in silico* identification of "optimal" drug targets. Their paper summarizes the need for a systems approach to understanding biology using methods developed in other disciplines. However, in contrast to many other engineering disciplines, the complexity of biology necessitates a reverse-engineering approach that requires the iteration and comparison of results obtained in computational modeling and wet lab experimentation.

Kirk E. Jordan
Emerging Solutions Executive
Systems and Technology Group

Steve Burbeck
Scientific Consultant