

Circulation Research

JOURNAL OF THE AMERICAN HEART ASSOCIATION

American Heart
Association® 
*Learn and Live*SM

Genome Informatics: Current Status and Future Prospects

Raimond L. Winslow and Mark S. Boguski

Circ. Res. 2003;92;953-961

DOI: 10.1161/01.RES.0000072475.04373.07

Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214

Copyright © 2003 American Heart Association. All rights reserved. Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://circres.ahajournals.org/cgi/content/full/92/9/953>

Subscriptions: Information about subscribing to Circulation Research is online at
<http://circres.ahajournals.org/subscriptions/>

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, 351 West Camden Street, Baltimore, MD 21202-2436. Phone 410-5280-4050. Fax: 410-528-8550. Email:
journalpermissions@lww.com

Reprints: Information about reprints can be found online at
<http://www.lww.com/static/html/reprints.html>

This Review is part of a thematic series on **Emerging Genomics Technology**, which includes the following articles:
DNA Microarrays: Implications for Cardiovascular Medicine
Serial Analysis of Gene Expression: Technical Considerations and Applications to Cardiovascular Biology

Genome Informatics: Current Status and Future Prospects

Technical Aspects of Screening for Genes or SNPs

Gordon F. Tomaselli, Editor

Genome Informatics Current Status and Future Prospects

Raimond L. Winslow, Mark S. Boguski

Abstract—This article reviews recent advances in genomics and informatics relevant to cardiovascular research. In particular, we review the status of (1) whole genome sequencing efforts in human, mouse, rat, zebrafish, and dog; (2) the development of data mining and analysis tools; (3) the launching of the National Heart, Lung, and Blood Institute Programs for Genomics Applications and Proteomics Initiative; (4) efforts to characterize the cardiac transcriptome and proteome; and (5) the current status of computational modeling of the cardiac myocyte. In each instance, we provide links to relevant sources of information on the World Wide Web and critical appraisals of the promises and the challenges of an expanding and diverse information landscape. (*Circ Res.* 2003;92:953-961.)

Key Words: cardiovascular genomics ■ transcriptome ■ proteome ■ modeling

The amount and rate of accumulation of biological information is increasing exponentially. This information explosion is being driven by development of powerful new technologies for acquiring large-scale genomic and proteomic datasets. Efforts are well underway to enumerate the cellular “parts lists” (genes and gene products) and to map and measure their dynamic interactions. The importance of these achievements cannot be understated, as they have transformed the nature of both biology and medicine, and have led many to claim that biology has become an information science.

These new experimental approaches are enabling acquisition of a wealth of new information on the cardiovascular genome, transcriptome, and proteome, information that will ultimately enhance our understanding of cardiovascular health and disease. At the same time, there is growing recognition that emergent, integrative behaviors of biological systems arise from the complex dynamic interactions be-

tween system components, and that knowledge of each component, however detailed, is not sufficient by itself to understand integrative behavior. A quantitative understanding of biological function will only be achieved through development of structurally, biochemically, and biophysically detailed computational models based directly on experimental data. Once developed, these models can be simulated, analyzed, and understood through application of modern engineering and computational approaches, and the knowledge gained from these analyses can be applied to the design of additional experiments. In this review, we attempt to identify and describe the model systems, information resources, and computational tools that are being developed to address the new research paradigm of systems biology.

The Genome Information Landscape

The amount of genome sequence data now available is enormous, with only more to come. At the time of this

Original received November 14, 2002; revision received March 17, 2003; accepted April 7, 2003.

From The Whitaker Biomedical Engineering Institute and Center for Cardiovascular Bioinformatics and Modeling (R.L.W.), The Johns Hopkins University School of Medicine and Whiting School of Engineering, Baltimore, Md; and the Human Biology Division (M.S.B.), Fred Hutchinson Cancer Research Center, Seattle, Wash.

This manuscript was sent to Richard A. Walsh, Consulting Editor, for review by expert referees, editorial decision, and final disposition.

Correspondence to Raimond L. Winslow, PhD, Room 201B Clark Hall, The Johns Hopkins University, 3400 N Charles St, Baltimore, MD 21218. E-mail rwinslow@bme.jhu.edu

© 2003 American Heart Association, Inc.

Circulation Research is available at <http://www.circresaha.org>

DOI: 10.1161/01.RES.0000072475.04373.07

TABLE 1. Genome Resources on the World Wide Web

| Description | URL |
|---|---|
| Guides to Finished Genomes and Current or Planned Genome Projects | |
| Genomes OnLine Database | http://ergo.integratedgenomics.com/GOLD/ |
| Entrez Genome | http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome |
| Genome News Network | http://gnn.tigr.org/sequenced_genomes/genome_guide_p1.shtml |
| NHGRI "White Paper" Proposals | http://www.genome.gov/page.cfm?pagelD=10002154 |
| Major Data Repositories and Information Resources | |
| US National Center for Biotechnology Information (NCBI) | http://www.ncbi.nlm.nih.gov/ |
| European Bioinformatics Institute (EBI) | http://www.ebi.ac.uk/Databases/index.html |
| Organism-Specific Databases | |
| Human NCBI | http://www.ncbi.nlm.nih.gov/genome/guide/human/ |
| EBI | http://www.ensembl.org/ |
| Santa Cruz | http://genome.ucsc.edu/ |
| Nature Genetics/NHGRI User's Guide | http://www.nature.com/cgi-taf/DynaPage.taf?file=/ng/journal/v32/n1s/index.html |
| Mouse | http://www.informatics.jax.org/ |
| Rat | http://rgd.mcw.edu/ |
| Zebrafish | http://zfin.org/cgi-bin/webdriver?Mval=aa-ZDB_home.apg |
| Dog | http://www.fhcrc.org/science/dog_genome/dog.html |

writing, the sequencing of more than 700 genomes is either finished or in progress, and results may be found in various public repositories (Table 1). The National Institutes of Health and six other US government agencies¹ are currently supporting genome projects and hundreds of researchers are submitting white paper proposals for the sequencing of additional organisms.² The following summary reviews those genome projects most pertinent to cardiovascular biology and medicine.

A "draft" sequence of the human genome has been available since June 2000, and the "finished" sequence was announced in April 2003, coinciding with the 50th anniversary of the determination of the double-helical structure of DNA.³ The first large-scale analyses of the 2.9-gigabase human genome sequence were published in February 2001.^{4,5} Of the many findings, the most surprising was that the human genome appears to contain roughly 30 000 protein-encoding genes, far fewer than the figure of 80 000 to 100 000 genes cited frequently in textbooks and only one and one-third to two times the number of genes in the fruit fly and the nematode worm. Comparative genomics studies have shown that many genes contributing to human disease are conserved among these genomes, underscoring the utility to biomedical research of studies in these organisms.^{6,7}

A draft sequence of the mouse genome and comparative analyses with the human sequence has been published.⁸ Findings support the notion that there are only about 30 000 genes in a typical mammalian genome. It is believed, however, that due to alternative splicing and other posttranscriptional and posttranslational modifications, this number of genes may encode a much larger number of functional proteins.^{9,10} Comparative human and mouse sequence analyses have also confirmed that, on average, rodent and human genes are about 85% identical in their coding sequences.^{11,12}

Although the mouse is the premier organism for studies of mammalian genetics and development, and is seeing in-

creased use for cardiovascular research (for example, by the JAX PGA and the Alliance for Cellular Signaling), the rat has been used more frequently for physiological and pharmacological studies.¹³ A consortium led by the National Heart, Lung, and Blood Institute (NHLBI) launched a rat genome program in 1995,¹⁴ which has, to date, produced a plethora of genomic resources,¹³ including genetic linkage maps that have been used to correlate genotypes with quantitative cardiovascular traits ("physiological profiles").¹⁵ Funding to sequence the rat genome was awarded in February 2001, and a draft sequence was released in November 2002.¹⁴

The zebrafish has become an important model organism for research in unraveling the molecular genetic basis of normal and abnormal cardiovascular form and function.¹⁶ Transparent embryos of this species have made possible large-scale screening of its genome for mutations with subsequent cloning of the affected genes.¹⁶ The availability of a genome sequence permits very rapid isolation of genes by "positional candidate" cloning¹⁷ once they are genetically mapped. The Sanger Center, funded by the Wellcome Trust, began zebrafish genome sequencing in February 2001.¹⁸ A very preliminary draft of the genome was released in July 2002.¹⁹

The dog has been a favored animal model for experimental medicine since the mid-19th century²⁰ and has been important in a wide variety of cardiovascular research applications, including the determination of potential cardiotoxicity of new drugs. Indeed, search of the CRISP database²¹ indicates that the canine model is used in hundreds of NHLBI grants. An impressive collection of canine genome resources (including physical, genetic, and transcript maps) is already available, and the dog was recently added to the "high priority" list of organisms for complete genomic sequencing.²²

Data Mining and Analysis

Since the advent of GenBank in 1982 and the Human Genome Project in 1990, bioinformatics has been synony-

mous with DNA and protein sequence data management and analysis,²³ and with the deluge of new data, this role for bioinformatics is unlikely to abate soon. Although bioinformatics has expanded in new directions with the emergence of functional genomics data and “systems biology,”^{24,25} the ability to navigate and search through sequence databases and associated annotations remains an essential skill. A number of books and reviews address the issues involved in sequence database searching.^{26,27} The following discussion is therefore limited to a few important characteristics of genome sequence data and annotation and to describing some typical analyses that are supported by existing databases and software tools. A case study of cardiovascular discovery through comparative genomics is also provided.

Genome sequence data are continually released into the public domain even before it has been finished, and thus, one must be mindful of this fact when analyzing data and interpreting results. Details of the production and assembly processes that result in draft versions of various levels of quality and completeness are discussed elsewhere.²⁸ Briefly, the accuracy of data are proportional to the number of times each nucleotide base in a particular sequence has been sampled, also known as “coverage.” For example, the preliminary draft of the zebrafish genome has only 2-fold coverage, whereas the rat genome is available at 6-fold coverage. Even 1-fold coverage data can be extremely useful for gene discovery, whereas 6-fold coverage data are of sufficient quality and continuity to justify and support detailed analysis and annotation.

Annotation of genome sequences is a complex process for which there is no real endpoint (see reference²⁹ for an excellent review). Consensus approaches involve similarity matching to previously sequenced cDNAs and/or genes, plus the application of various gene prediction programs, but the major information providers (Table 1) differ in details and emphasis.²⁸ Annotation “pipelines,” and the reliability of the resultant products, continue to evolve, and scientific interpretation of the genome will perpetually be subject to revisions as new discoveries are made. To increase the utility of these data for the nonspecialist, the National Human Genome Research Institute (NHGRI), in collaboration with the journal *Nature Genetics*, has produced “A User’s Guide to the Human Genome.”²⁸ This tutorial consists of 13 Web-based exercises that illustrate how to solve a variety of common, but powerful and sophisticated tasks, such as in silico positional cloning, the analysis of gene families, and the identification of functional and structural domains in proteins.

To illustrate the power of comparative genomics, we cite the recent discovery of the apolipoprotein AV gene (*APOAV*)³⁰ that encodes a previously unknown member of the well-studied apolipoprotein gene family, mostly closely related to *APOAIV*, which was cloned nearly two decades ago.³¹ The *APOAIV*, *APOCIII*, and *APOAI* genes are found within a 20-kb locus on human chromosome 11q32.³² Through comparative analysis of human and mouse genomic sequences, Pennacchio et al³⁰ discovered a region of sequence conservation approximately 25 kb downstream from *APOAIV* that proved to contain the *APOAV* gene. Because the *AICIII/AIV* locus was well-known to influence plasma lipid levels in

humans, Pennacchio et al studied lipid levels in knockout and transgenic mice and showed that *APOAV* has a strong inverse correlation with plasma triglyceride levels.³⁰ Subsequent studies by this group demonstrated that genetic polymorphisms in the *APOAV* locus are significantly associated with plasma triglyceride levels in humans.

The Cardiovascular Transcriptome

Identification of the cardiac transcriptome is a critically important first step toward understanding how environmental factors and disease processes affect gene expression in the heart. No consensus cardiac transcriptome has yet been established for any organism, including human. Indeed, there are few publicly available resources describing gene expression in heart. Significance of the problem is well illustrated using the example of gene *KCND3*. *KCND3* is known to be expressed in the cardiac ventricles and is thought to encode the major component of the voltage-dependent Ca^{2+} -independent transient outward current (I_{to1}), a key contributor in shaping the early phase of the cardiac ventricular action potential.³³ *KCND3* mRNA transcript level is also known to be downregulated in end-stage heart failure.³⁴ A search of GenBank using the string “Kv4.3 OR *KCND3*” indicates a total of 75 entries, 25 of which are from human and 8 of which have the designation “heart” in the tissue type field. *KCND3* corresponds to Unigene cluster Hs.184889 and Locus Link id 3752. However, the heart tissue type annotation from the 8 GenBank entries is not carried forward to the Unigene cluster, and Locus Link reports do not contain a tissue type field. Indeed, the Unigene cluster summary indicates in the expression information field that *KCND3* is expressed in neural tissue. This illustrates the difficulty of identifying genes expressed in heart.

There have been limited efforts to organize information on the cardiac transcriptome. The BodyMap Database^{35,36} describes tissue-specific gene expression, including that in human left ventricle, right atria, and embryonic mouse heart. To obtain this database, a 3'-directed cDNA library was prepared from each tissue sample, and randomly selected clones were sequenced, compared, and organized into clusters. A representative sequence having the lowest content of ambiguous bases was selected from each cluster and compared against data in GenBank. Those with over 90% similarity to the 3' end of the mRNA entries or to the reported terminal exon of known genes were regarded as representing those genes. BodyMap lists 744 genes expressed in human heart (atria and ventricles) and 453 genes expressed in embryonic mouse heart. Updates of BodyMap entries are infrequent and database query capabilities are limited.

An additional resource is the Cardiac Gene Expression (CaGE) Knowledgebase.^{37,38} In order for a gene to be included in CaGE, it must be present in NCBI’s Locus Link^{39–41} and have evidence confirming its expression in heart. Supporting evidence includes the designation of heart in the express field of a Unigene cluster, or the designation of heart in the tissue type field of a GenBank⁴² entry. Relationships can then be established between any individual GenBank clone assigned to a Unigene cluster and any Unigene cluster representing a unique Locus Link. Two additional

sources of human cardiac gene expression data accessed by CaGE are the Toronto Cardiac Gene Unit library⁴³ and the BodyMap database.³⁶ CaGE is rebuilt daily after accessing all of these sources. The last source of evidence for expression comes from gene expression profiling experiments conducted on normal and dilated cardiomyopathic failing human hearts. Currently, such data are limited to that collected at the Johns Hopkins University School of Medicine using oligonucleotide and cDNA microarrays.⁴⁴ Plans are being developed to incorporate data from expression studies conducted as part of the NHLBI Programs in Genomic Applications, described later.

Interaction with CaGE is via a Web interface. Genes can be browsed by the first letter or number of their official gene name. Basic searches can be performed using either official or alias gene names, chromosome number or cytogenetic band, official or alias gene symbols, GenBank accession number, Unigene cluster, or Locus Link identifiers. Clinical synopses found within OMIM⁴⁵ may be searched for genes known to be associated with a given human phenotype. The results of these queries generate a gene "home page" that displays all data stored within CaGE for that given locus link. Currently, CaGE contains 7349 Unigene clusters known to be expressed in human cardiac tissue. There are 676 human Locus Link entries in the BodyMap Atria library, 721 in the BodyMap Ventricle library, and 2618 in the Toronto Cardiac Gene Unit library. An additional 1800 human Locus Links are from the Johns Hopkins gene expression data. In total, CaGE tabulates 8085 unique Human Locus Link entries expressed in human cardiac tissue. This is just a fraction of the total number of genes estimated to be expressed in aorta, adult, and fetal heart.⁴⁶

Serial analysis of gene expression (SAGE) is a powerful method for the identification of gene expression patterns.^{47,48} Advantages over other methods such as use of oligo- or cDNA microarrays are that SAGE is not dependent on prior knowledge of transcript information and is able to detect transcripts expressed at low copy number. Results are reported in terms of absolute or relative numbers of tags, facilitating direct comparison of SAGE results obtained in different laboratories. NCBI has developed a public repository for SAGE transcriptome information from a number of different organisms and tissues called SAGEmap.^{49,50} Recently, Anisimov et al⁵¹ have used SAGE to produce the first quantitative expression profile of adult mouse heart and have made this transcriptome available at SAGEmap (GSM1681). This represents an important step forward in the quantitative determination of the cardiac transcriptome and is an approach that is likely to be extended to other species in the near future.

One of the most exciting new experimental technologies to emerge in recent years have been methods for obtaining genome-wide mRNA expression data using oligonucleotide⁵² and cDNA microarrays,⁵³ a topic considered previously in this Review series.⁵⁴ These approaches are likely to provide significant insights into changes of gene expression as well as mechanisms of gene regulation in a variety of cardiac disease processes.⁵⁵⁻⁶⁰ The National Heart, Lung, and Blood Institute (NHLBI) launched the Programs for Genomic Applications (PGAs)⁶¹ on September 30, 2000, funding a total of 11

projects (Table 2). This program is a major initiative to advance functional genomic research relating to heart, lung, blood, and sleep health and disorders, with the majority of sites undertaking significant research efforts in large-scale studies of gene expression in human cardiovascular disease as well as in animal models. Specific goals of the PGAs include the following: (1) development of animal models and characterization of phenotype in these models; (2) measurement of gene expression, identification of regulated genes, and identification of single nucleotide polymorphisms (SNPs) in both animal models and human patients for a range of cardiopulmonary disorders; (3) development of new databases, data analysis procedures, and software tools for cardiovascular genomics. All information, reagents, and tools developed by the PGAs are mandated to be released in a timely manner to the research community. Microarray data are released 60 days after completion of the last hybridization of an individual experiment, and after the data passes quality control standards. Data are made available either in the form of text files (the Hopgenes PGA) or through database query (eg, chipperDB of the CardioGenomics PGA or GeneTraffic 2.1 of the Southwestern PGA). Distribution of microarray data will be facilitated greatly by adoption of MIAME (Minimal Information for Annotation of Microarray Experiments)⁶²-compliant data markup languages such as the Microarray Gene Expression Markup Language (MAGE-ML).⁶³ Indeed, the *Nature* journals and *The Lancet* have announced recently that microarray data submitted for publication must be MIAME-compliant.

The Cardiovascular Proteome

Proteomics is the study of the full protein complement of the genome, and seeks to identify and characterize mechanisms regulating expression level, co- and posttranslational modifications, and interactions between all proteins in the cell. Applications of proteomics to the study of cardiovascular biology are described in several recent reviews.⁶⁴⁻⁶⁶ These efforts are presently focused on characterization of the cardiac cytoplasmic, mitochondrial, and myofilament subproteomes⁶⁷⁻⁶⁹ and characterization of myocyte response to ischemic injury.⁷⁰⁻⁷⁴ Research in cardiovascular proteomics is likely to develop rapidly, particularly in light of the recent funding of ten national proteomics centers as part of the NHLBI Proteomics Initiative. Similar in spirit to the Programs for Genomic Applications, the intent of this initiative is to establish a number of multidisciplinary centers focused on the development of innovative proteomics technologies, and the application of these technologies to enhancing our understanding of heart, lung, blood, and/or sleep disease. As with the PGA, all products of this effort including reagents, experimental and analytical techniques, and data will be made available to the scientific community. This initiative is sufficiently new that at the time of this writing, URLs for center Web sites are not yet available.

There are a number of well-known protein databases that make available summary information on general protein sequence and structure. Information in these databases is typically generated by expert annotators. Examples include the SWISS-PROT/TrEMBL Protein Knowledgebase,⁷⁵⁻⁷⁷ the Protein Infor-

TABLE 2. NHLBI Programs in Genomics Applications

| Name | Project Title | Project Web Site | Resources |
|----------------|---|---|---|
| BayGenomics | The NHLBI-Bay Area Functional Genomics Consortium | http://baygenomics.ucsf.edu/ | Murine embryonic stem cell lines for use in cardiovascular and pulmonary research |
| Berkeley PGA | Comparative Genomic Analysis of Cardiovascular Gene Regulation | http://pga.lbl.gov | VISTA tools for comparative genomics; CVCGD database of cardiovascular genes; NOMAD microarray database |
| CardioGenomics | Genomics of Cardiovascular Development, Adaptation and Remodeling | http://www.cardiogenomics.org | Expression profiles from murine models of cardiomyopathy; sarcomere protein gene mutation database; human gene mutation and SNP data |
| HopGenes | Applied Genomics in Cardiopulmonary Disease | http://www.hopkins-genomics.org | affyNorm A microarray analysis software; cardiopulmonary disease candidate genes list; expression profiles from rat and murine models of pulmonary disease |
| InnateImmunity | Innate Immunity in Heart, Lung, and Blood Disease | http://innateimmunity.net | Innate immunity gene list; SNP data analysis tools |
| JAX PGA | Mouse Models of Heart, Lung, and Blood Diseases | http://pga.jax.org | Murine models of heart, lung, blood, and sleep disorders, mouse phenome database |
| ParaBioSys | Genomic Analysis of Stress and Inflammation | http://pga.mgh.harvard.edu | ParaBioSys software tools for genomic searching and alignment; murine expression data in response to proinflammatory, metabolic, and pathogen stresses affecting the cardiovascular system and the lung |
| PhysGen | Physiogenomics of Stressors in Derived Consomic Rats | http://pga.mcw.edu | Consomic rat strains for investigating physiological mechanisms of homeostasis and response to stress; phenotypic, genotypic, and microarray data |
| Seattle SNPs | UW-FHCRC Variation Discovery Resource | http://pga.mbt.washington.edu | Human inflammatory response SNP database; PolyPhred SNP detection software; Genotype (VG2) and Haplotype (VH1) visualization software |
| Southwestern | Genomics and Proteomics of Cell Injury and Inflammation | http://pga.swmed.edu | Expression and SNP data on genes involved in the cellular response to injury an inflammation; GeneTraffic microarray database |
| TREX | Microarray Expression Profiling of Rodent Models of Human Disease | http://pga.tigr.org | TIGR Microarray Data Analysis System; expression profiles for murine and rat models of heart, lung, and blood disease |

mation Resource,^{78,79} and the Protein Data Bank.^{80,81} NCBI's *Entrez-Protein* (www.ncbi.nlm.nih.gov/Database/index.html) compiles and makes available information from many of these sources; information retrieval is via Boolean keyword search. A variety of other data analysis and retrieval tools, such as the BLAST program for sequence similarity searching, are also available (www.ncbi.nlm.nih.gov/Tools/index.html).

Databases supporting direct access to experimental data currently take the form of annotated flat-file images of 2-D gels. These databases include SWISS-2DPAGE as well as heart-specific 2-D gel databases (the Harefield Hospital Heart Science Center 2D Gel Protein Database HSC-2D PAGE⁸²; the Heart High-Performance 2-DE Database at the Max Delbruck Center for Molecular Medicine⁸³; the Human Myocardial Two-Dimensional Electrophoresis Protein Database Heart 2D-PAGE⁸⁴; and the 2-dimensional polyacrylamide gel electrophoresis database of rat heart⁸⁵). Creation of such databases is facilitated by the availability of open source software called "make2ddb" for generating 2-D gel databases.⁸⁶

These data sources constitute a rich resource for the general proteomics community. However, the rapid pace of development of proteomics technologies and the resulting diversity and complexity of proteomics data poses special challenges.⁸⁷ In particular, methods for structuring and searching proteomics databases to retrieve groups of proteins based on well-known pathways, functional classifications,

and specific posttranslational modifications must be developed. Methods for annotating and differentiating posttranslational modifications predicted from protein motifs using computational algorithms versus those for which there is direct experimental evidence are required. Protein concentrations and other measured attributes should be compared with values determined in reference samples to enhance data quantification. With regard to protein identification based on mass spectrometry, annotations must provide meaningful statistical measures of the quality of match. An issue that will figure importantly in the development of the NIH Proteomics Centers is that data representation and dissemination must be facilitated by the adoption of standards for data description. As one example of such an effort, the Human Proteome Organization is promoting the development of standard formats for the representation and exchange of mass spectrometry and protein-protein interaction data and annotations.^{88,89} These formats are derivatives of XML (eXtensible Markup Language), a language that originated as a standard for document formatting, but which is now used as a format to transfer structured data of any kind over the World Wide Web. Finally, Web services,⁹⁰ a technology building on the ability of Simple Object Access Protocol (SOAP) to support distributed network communication, has great potential as a tool for making both data and computational algorithms transparently available to other software applications, thus

facilitating the machine discovery, communication, and analyses of proteomic as well as genomic data.

The data resources described above provide descriptions of the properties of individual genes and proteins. However, cellular behavior is regulated in a complex manner through a diversity of interacting gene expression, signal transduction, metabolic, and electrophysiological pathways. Pathway properties are themselves determined by factors such as the specific nature of molecular interactions, formation of multimolecular complexes, and by subcellular localization. Representation of information on biological pathways in a form that supports complex querying and modeling is an important goal of postgenomics biology.

There are emerging databases and bioinformatics tools that address this need. The most ambitious effort is that of The Alliance for Cellular Signaling (AFCS). The aim of the AFCS is to achieve a quantitative understanding of cellular G protein-mediated signaling. This will be done using the resting B lymphocyte and the murine cardiac myocyte as model systems. Goals are to identify all of the proteins comprising signaling pathways within these cells, to determine how interactions between these proteins determine spatiotemporal responses within the cell, and to assemble from these data theoretical and computational models of cellular signaling. The AFCS is developing an object-relational database of information on signaling proteins called Molecule Pages. This database is accessible through both the AFCS Web site⁹¹ and the AFCS/Nature Signaling Gateway.⁹² This database consists of annotations extracted automatically from public data repositories on the function of over 3000 proteins involved in signal transduction. In addition, over 800 so-called "Mini-Molecule Pages" consisting of expert protein annotations are available. These experts are responsible for maintaining annotation information as the project progresses. Ultimately, the AFCS database will be extended to support queries regarding relationships between molecules and the structure of complex signaling pathways.

There are additional software resources that, although not cardiac-specific, could prove to be of great value to the cardiovascular community. One of these is the Bimolecular Interaction Network Database (BIND).⁹³ BIND is an object-relational database for storing and querying information on pairwise protein-protein interactions, protein complexes, and protein pathways. The BIND data model has been published^{94,95} and is sufficiently rich that it permits description of interactions between proteins, nucleic acids, and small molecules, protein post-translational modifications, linkage to external data sources for annotation purposes, and organization of pairwise interactions into larger scale interaction networks. A BIND Interaction Viewer is available for visualization of network interactions. BIND is an Open Source software development project and is available for download from SourceFORGE.net. Both BIND and the AFCS Molecules Pages therefore represent important resources that could be extended and used for development of cardiac-specific protein interaction and pathway databases.

Gene MicroArray Pathway Profiler^{96,97} is a free software tool for viewing and analyzing gene expression data superimposed on drawings of gene interaction networks with

hyperlinks to annotation data. The software consists of several components. The GenMAPP Drafting Board software, in conjunction with Drafting Tools and the Object Toolbox, enable users to create new pathway representations as well as edit existing pathways, and to store these pathways in a file format known as MAPPs. The Expression Data Manager software is used to import and layer expression data onto these MAPPs. The GenMAPP Database is a library of all the genes used by the GenMAPP software. The GenMAPP Database stores information for linking expression data to objects in a MAPP, and also stores annotations for each MAPP object.

The PathDB software suite from National Center for Genome Research⁹⁸ is an alternative system for building, visualizing, and querying cellular networks. Using client-side components (QueryTool and PathwayViewer), users may view and query pathway models stored in the PathDB relational database. Users may also download the software suite in order to run it locally to create and store pathway models.

New Directions in Bioinformatics: Computational Modeling

Integrative modeling of the cardiac myocyte has been advanced to a greater breadth and depth than that achieved in any other discipline of biological modeling. Development of myocyte models began in the early 1960s with publication of Purkinje fiber action potential models based on the Hodgkin-Huxley model of the squid action potential.^{99,100} Subsequent elaboration of these and other models led to development of the first biophysically based cell model describing interactions between voltage-gated membrane currents, pumps and exchangers, and intracellular calcium cycling processes in the cardiac myocyte,¹⁰¹ the so-called DiFrancesco-Noble model of the Purkinje fiber. This landmark model established the conceptual framework from which all subsequent models of the myocyte have been derived. Models of the myocyte now include descriptions of (1) voltage-dependent membrane currents, in some instances based on formulation of Markov state models of ion channels¹⁰²⁻¹⁰⁴; (2) membrane pump and transporter function; (3) intracellular calcium cycling^{104,105}; (4) excitation-contraction coupling and isometric force generation¹⁰⁶; and (5) energy production via the tricarboxylic acid cycle and oxidative phosphorylation.¹⁰⁷⁻¹⁰⁹ These models have proven reproductive and predictive properties and have been applied to advance our understanding of myocyte function in both health and disease.^{102,110-112}

Source code is now available for at least three models of the ventricular myocyte action potential (the Luo-Rudy Dynamic model of the mammalian action potential,¹¹³ the Winslow-Rice-Jafri model of the canine ventricular myocyte, and the Jafri-Rice-Winslow model of the guinea pig ventricular action potential¹¹⁴). In addition, there are Web-based simulation resources that facilitate model dissemination and use. These include the following: (1) the Virtual Cell system of the National Resource for Cell Analysis and Modeling at the University of Connecticut¹¹⁵; (2) the Java-Based Integrative Model Simulation and Analysis Environment (JSIM)¹¹⁶—an Open Source software devel-

opment project of the National Simulation Resource at University of Washington, Seattle; and (3) iCell¹¹⁷—a collection of myocyte models implemented as Java applets.

The complexity of biological models, including those of the cardiac myocyte, is increasing rapidly. This complexity makes the reliable publication and exchange of models difficult. XML-based markup languages such as CellML¹¹⁸ and the Systems Biology Markup Language (SBML)¹¹⁹ are being developed to support the error-free exchange of models independently of the hardware and software architectures on which these models will run. An application programming interface for CellML is being developed, and several groups are developing software for automated source code generation from CellML files.

Future Directions

A national infrastructure supporting the acquisition, distribution, and analysis of cardiovascular genomic and proteomic data is now in the formative stage. We will, without question, witness dramatic growth of the quantity, quality, and availability of cardiovascular data and models relating to health and disease over the next five years. However, the value of these data and models will depend to a great extent on quality of the annotation provided. In the case of experimental data, it is necessary that investigators undertake critical review and quality control before public release, and provide careful annotation to assure that animal model phenotype, clinical information relevant to the interpretation of samples from human tissue, and all aspects of sample preparation and analysis are described fully. Data must be organized within databases having a sufficiently rich schema to permit complex queries based on underlying data attributes. Annotations describing data processing methods applied to any archived data must be available. In the case of computational models, annotation must include documentation as to how model parameters are determined, and evidence of model reproductive and predictive capabilities. Finally, data and models must be exportable/accessible using standards agreed on by the research community so as to facilitate error-free machine exchange. If these challenges are met, we will have the opportunity to create a truly integrated cardiovascular research community, the whole of which is far greater than the sum of its parts.

Acknowledgments

This work is supported by grants from the NIH (RO1 HL-61711, RO1 HL-60133, RO1 HL-72488, P50 HL-52307, N01 HV-28180), The Falk Medical Trust, and The Whitaker Foundation.

References

1. National Human Genome Research Institute. Online research resources: other federal projects in genomics. Available at: <http://www.genome.gov/page.cfm?pageID=10003899>
2. National Human Genome Research Institute. Sequences, maps and BAC libraries: genome sequencing prioritization list. Available at: <http://www.genome.gov/page.cfm?pageID=10002154>
3. Watson JD, Crick FHC. A structure for deoxyribose nucleic acid. *Nature*. 1953;171:737.
4. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
5. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. The sequence of the human genome. *Science*. 2001;291:1304–1351.
6. Wheelan SJ, Boguski MS, Duret L, Makalowski W. Human and nematode orthologs—lessons from the analysis of 1800 human genes and the proteome of *Caenorhabditis elegans*. *Gene*. 1999;238:163–170.
7. Fortini ME, Skupski MP, Boguski MS, Hariharan IK. A survey of human disease gene counterparts in the *Drosophila* genome. *J Cell Biol*. 2000;150:F23–F30.
8. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 2002;420:520–562.
9. Maniatis T, Tasic B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature*. 2002;418:236–243.
10. Roberts GC, Smith CW. Alternative splicing: combinatorial output from the genome. *Curr Opin Chem Biol*. 2002;6:375–383.
11. Makalowski W, Zhang J, Boguski MS. Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. *Genome Res*. 1996;6:846–857.
12. Makalowski W, Boguski MS. Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc Natl Acad Sci U S A*. 1998;95:9407–9412.
13. Jacob HJ, Kwitek AE. Rat genetics: attaching physiology and pharmacology to the genome. *Nat Rev Genet*. 2002;3:33–42.
14. NIH Rat Genomics and Genetics Web site. Available at: <http://www.nih.gov/science/models/rat/>
15. Stoll M, Cowley AW Jr, Tonellato PJ, Greene AS, Kaldunski ML, Roman RJ, Dumas P, Schork NJ, Wang Z, Jacob HJ. A genomic-systems biology map for cardiovascular function. *Science*. 2001;294:1723–1726.
16. Shin JT, Fishman MC. From zebrafish to human: modular medical models. *Annu Rev Genomics Hum Genet*. 2002;3:311–340.
17. Collins FS. Positional cloning moves from perdito to traditional. *Nat Genet*. 1995;9:347–350.
18. The Wellcome Trust Sanger Institute. The *Danio rerio* sequencing project. Available at: http://www.sanger.ac.uk/Projects/D_rerio/
19. The Wellcome Trust Sanger Institute. First assembly of the zebrafish genome released: The *Danio rerio* sequencing project. Available at: http://www.sanger.ac.uk/Projects/D_rerio/assembly_information.shtml
20. Bernard C. *An Introduction to the Study of Experimental Medicine*. New York, NY: Dover; 1957.
21. ERA Commons: Computer Retrieval of Information on Scientific Projects. Available at: <http://crisp.cit.nih.gov/>
22. National Human Genome Research Institute. Sequences, Maps and BAC Libraries: Genome Sequencing Prioritization List. Available at: <http://www.genome.gov/page.cfm?pageID=10002154>
23. Boguski MS. Biosequence exegesis. *Science*. 1999;286:453–455.
24. Hieter P, Boguski M. Functional genomics: it's all how you read it. *Science*. 1997;278:601–602.
25. Kitano H. Systems biology: a brief overview. *Science*. 2002;295:1662–1664.
26. Altschul SF, Boguski MS, Gish W, Wootton JC. Issues in searching molecular sequence databases. *Nat Genet*. 1994;6:119–129.
27. Pickeral O, Boguski MS. The bioinformatics bookshelf: teach yourself computational biology. *Cell*. 1999;96:451–455.
28. Wolfsberg TG, Wetterstrand KA, Guyer MS, Collins FS, Baxevanis AD. A user's guide to the human genome. *Nat Genet*. 2002;32(suppl):1–79.
29. Stein L. Genome annotation: from sequence to biology. *Nat Rev Genet*. 2001;2:493–503.
30. Pennacchio LA, Olivier M, Hubacek JA, Cohen JC, Cox DR, Fruchart JC, Krauss RM, Rubin EM. An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science*. 2001;294:169–173.
31. Boguski MS, Elshourbagy N, Taylor JM, Gordon JI. Rat apolipoprotein A-IV contains 13 tandem repetitions of a 22-amino acid segment with amphipathic helical potential. *Proc Natl Acad Sci U S A*. 1984;81:5021–5025.
32. Elshourbagy NA, Walker DW, Boguski MS, Gordon JI, Taylor JM. The nucleotide and derived amino acid sequence of human apolipoprotein A-IV mRNA and the close linkage of its gene to the genes of apolipoproteins A-I and C-III. *J Biol Chem*. 1986;261:1998–2002.
33. Li G-R, Feng J, Yue L, Carrier M. Transmural heterogeneity of action potentials and Ito1 in myocytes isolated from the human right ventricle. *Am J Physiol*. 1998;275:H369–H377.

34. Kaab S, Dixon J, Duc J, Ashen D, Nabauer M, Beuckelmann DJ, Steinbeck D, McKinnon D, Tomaselli GF. Molecular basis of transient outward potassium current downregulation in human heart failure: a decrease in Kv4.3 mRNA correlates with a reduction in current density. *Circulation*. 1998;98:1383–1393.
35. Human and Mouse Gene Expression Database. BodyMap. Available at: <http://bodymap.ims.u-tokyo.ac.jp/>.
36. Okubo K, Hori N, Matoba R, Niiyama T, Fukushima A, Kojima Y, Matsubara K. Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nat Genet*. 1992;2:173–179.
37. Bober M, Wiehe K, Yung C, Onal Suzek T, Lin M, Baumgartner W Jr, Winslow R. CaGE: cardiac gene expression knowledgebase. *Bioinformatics*. 2002;18:1013–1014.
38. The Cardiac Gene Expression Knowledgebase. Available at: <http://www.cage.wbmei.jhu.edu>
39. Pruitt KD, Katz KS, Sicotte H, Maglotte DR. Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. *Trends Genet*. 2000;16:44–47.
40. Pruitt KD, Maglott DR. RefSeq, and LocusLink. NCBI gene-centered resources. *Nucleic Acids Res*. 2001;29:137–140.
41. National Center for Biotechnology Information. Locus Link. Available at: <http://www.ncbi.nlm.nih.gov/LocusLink/>
42. Benson DA, Boguski MS, Lipman DJ, Ostell J, Ouellette BF, Rapp BA, Wheeler DL. GenBank. *Nucleic Acids Res*. 1999;27:12–17.
43. Hwang DM, Dempsey AA, Wang RX, Rezvani M, Barrans JD, Dai KS, Wang HY, Ma H, Cukerman E, Liu YQ, Gu JR. A genome-based resource for molecular cardiovascular medicine: toward a compendium of cardiovascular genes. *Circulation*. 1997;96:4146–4203.
44. Bober MB, Delmar P, Szak S, Karaoz U, White JA, Hudson J, Boguski M, Tomaselli GF, Winslow RL. Gene expression in human heart failure: microarray analysis of dilated cardiomyopathy. In: *Currents in Computational Molecular Biology*. Montreal, Canada: Les Publications CRM; 2001:171–172.
45. OMIM (Online Mendelian Inheritance in Man). McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University (Baltimore, Md) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md); 2000. Available at: <http://www.ncbi.nlm.nih.gov/omim/>
46. Dempsey AA, Dzau VJ, Liew CC. Cardiovascular genomics: estimating the total number of genes expressed in the human cardiovascular system. *J Mol Cell Cardiol*. 2001;33:1879–1886.
47. Velculescu VE, Vogelstein B, Kinzler KW. Analysing uncharted transcriptomes with SAGE. *Trends Genet*. 2000;16:423–425.
48. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science*. 1995;270:484–487.
49. Lash AE, Tolstoshev CM, Wagner L, Schuler GD, Strausberg RL, Riggins GJ, Altschul SF. SAGEmap. a public gene expression resource. *Genome Res*. 2000;10:1051–1060.
50. Serial Analysis of Gene Expression. SAGEmap Web site. Available at: <http://www.ncbi.nlm.nih.gov/sage>
51. Anisimov SV, Tarasov KV, Stern MD, Lakatta EG, Boheler KR. A quantitative and validated SAGE transcriptome reference for adult mouse heart. *Genomics*. 2002;80:213–222.
52. Lipshutz RJ, Morris D, Chee M, Hubbell E, Kozal MJ, Shah N, Shen N, Yang R, Fodor SP. Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques*. 1995;19:442–447.
53. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995;270:467–470.
54. Cook SA, Rosenzweig A. DNA microarrays: implications for cardiovascular medicine. *Circ Res*. 2002;91:559–564.
55. Barrans JD, Stamatou D, Liew C. Construction of a human cardiovascular cDNA microarray: portrait of the failing heart. *Biochem Biophys Res Commun*. 2001;280:964–969.
56. Barrans JD, Allen PD, Stamatou D, Dzau VJ, Liew CC. Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. *Am J Pathol*. 2002;160:2035–2043.
57. Stanton LW, Garrard LJ, Damm D, Garrick BL, Lam A, Kapoun AM, Zheng Q, Protter AA, Schreiner GF, White RT. Altered patterns of gene expression in response to myocardial infarction. *Circ Res*. 2000;86:939–945.
58. Yabg J, Moravec CS, Sussman MA, DiPaola NR, Hawthorne L, Mitchell CA, Young JB, Francis GS, McCarthy PM, Bond M. Decreased SLIM1 expression and increased gelsolin expression in failing human hearts measured by high-density oligonucleotide arrays. *Circulation*. 2000;102:3046–3052.
59. Zhu Y, Yang H-T, Boheler KR. Identification of novel transcripts implicated in the progression to a senescent myocardium: results from microarrays. *Circulation*. 2000;102(suppl II):II-142. Abstract.
60. Tan F, Moravec C, Li J, Apperson-Hansen C, McCarthy P, Young J, Bond M. The gene expression fingerprint of human heart failure. *Proc Natl Acad Sci U S A*. 2002;99:11387–11392.
61. National Heart, Lung, and Blood Institute. Programs for Genomic Applications Home Page. Available at: <http://www.nhlbi.nih.gov/resources/pga/index.htm>
62. Brazza A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoekert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*. 2001;29:365–371.
63. Spellman PT, Miller M, Stewart J, Troup C, Sarkans U, Chervitz S, Bernhart D, Sherlock G, Ball C, Lepage M, Swiatek M, Marks WL, Goncalves J, Markel S, Jordan D, Shojatalab M, Pizarro A, White J, Hubley R, Deutsch E, Senger M, Aronow BJ, Robinson A, Bassett D, Stoekert CJ Jr, Brazza A. Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol*. 2002;3:RESEARCH0046.
64. Arrell DK, Neverova I, Van Eyk JE. Cardiovascular proteomics: evolution and potential. *Circ Res*. 2001;88:763–773.
65. Macri J, Rapundalo ST. Application of proteomics to the study of cardiovascular biology. *Trends Cardiovasc Med*. 2001;11:66–75.
66. Van Eyk JE. Proteomics: unraveling the complexity of heart disease and striving to change cardiology. *Curr Opin Mol Ther*. 2001;3:546–553.
67. Neverova I, Van Eyk JE. Application of reversed phase high performance liquid chromatography for subproteomic analysis of cardiac muscle. *Proteomics*. 2002;2:22–31.
68. Labugger R, McDonough JL, Neverova I, Van Eyk JE. Solubilization, two-dimensional separation and detection of the cardiac myofibrillar protein troponin T. *Proteomics*. 2002;2:673–678.
69. Lopez MF, Melov S. Applied proteomics: mitochondrial proteins and effect on function. *Circ Res*. 2002;90:380–389.
70. Vondriska TM, Klein JB, Ping P. Use of functional proteomics to investigate PKC epsilon-mediated cardioprotection: the signaling molecule hypothesis. *Am J Physiol Heart Circ Physiol*. 2001;280:H1434–H1441.
71. Vondriska TM, Ping P. Functional proteomics to study protection of the ischaemic myocardium. *Expert Opin Ther Targets*. 2002;6:563–570.
72. McDonough JL, Arrell DK, Van Eyk JE. Troponin I degradation and covalent complex formation accompanies myocardial ischemia/reperfusion injury. *Circ Res*. 1999;84:9–20.
73. Arrell DK, Neverova I, Fraser H, Marban E, Van Eyk JE. Proteomic analysis of pharmacologically preconditioned cardiomyocytes reveals novel phosphorylation of myosin light chain 1. *Circ Res*. 2001;89:480–487.
74. Edmondson RD, Vondriska TM, Biederman KJ, Zhang J, Jones RC, Zheng Y, Allen DL, Xiu JX, Cardwell EM, Pisano MR, Ping P. Protein kinase C epsilon signaling complexes include metabolism- and transcription/translation-related proteins: complimentary separation techniques with LC/MS/MS. *Mol Cell Proteomics*. 2002;1:421–433.
75. Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res*. 1999;27:49–54.
76. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*. 2000;28:45–48.
77. Gasteiger E, Jung E, Bairoch A. SWISS-PROT: connecting biomolecular knowledge via a protein database. *Curr Issues Mol Biol*. 2001;3:47–55.
78. Wu CH, Huang H, Arminski L, Castro-Alvarez J, Chen Y, Hu ZZ, Ledley RS, Lewis KC, Mewes HW, Orcutt BC, Suzek BE, Tsugita A, Vinayaka CR, Yeh LS, Zhang J, Barker WC. The Protein Information Resource: an integrated public resource of functional annotation of proteins. *Nucleic Acids Res*. 2002;30:35–37.
79. Protein Information Resource. Available at: <http://pir.georgetown.edu/>
80. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD,

- Zardecki C. The protein data bank. *Acta Crystallogr D Biol Crystallogr*. 2002;58:899–907.
81. Protein Data Bank. Available at: <http://www.rcsb.org/pdb/>
 82. HSC-2DPAGE. 2-DE Gel Protein Databases at Harefield. Available at: <http://www.harefield.nthames.nhs.uk/nhli/protein/>
 83. The Max Delbrück Center for Molecular Medicine. Heart High-Performance 2-DE Database. Available at: <http://www.mdc-berlin.de/~emu/heart/heart.html>
 84. HEART-2DPAGE. The Human Myocardial Two-Dimensional Electrophoresis Protein Database. Available at: <http://userpage.chemie.fu-berlin.de/~pleiss/dhzb.html>
 85. RAT HEART-2DPAGE. Two-dimensional polyacrylamide gel electrophoresis database of rat heart. Available at: <http://www.mpiib-berlin.mpg.de/2D-PAGE/RAT-HEART/2d/>
 86. ExpASY Home Page. Request for Make2ddb package. Available at: <http://www.expasy.org/ch2d/make2ddb.html>
 87. Boguski MS, McIntosh MW. Biomedical informatics for proteomics. *Nature*. 2003;422:233–237.
 88. Orchard S, Kersey P, Hermjakob H, Apweiler R. The HUPO proteomics standards initiative meeting: towards common standards for exchanging proteomics data. *Comp Funct Genomics*. 2003;4:16–19.
 89. Taylor CF, Paton NW, Garwood KL, Kirby PD, Stead DA, Yin Z, Deutsch EW, Selway L, Walker J, Riba-Garcia I, Mohammed S, Deery MJ, Howard JA, Dunkley T, Aebersold R, Kell DB, Lilley KS, Roepstorff P, Yates JR, Brass A, Brown AJ, Cash P, Gaskell SJ, Hubbard SJ, Oliver SG. A systematic approach to modeling, capturing, and disseminating proteomics experimental data. *Nat Biotechnol*. 2003; 21:247–254.
 90. Colye FP. *XML, Web Services, and the Data Revolution*. Boston, Mass: Addison-Wesley; 2002.
 91. Alliance for Cellular Signaling Web site. Available at: <http://www.afcs.org>
 92. AFCS Nature. A comprehensive signaling database. Available at: <http://www.signaling-gateway.org/molecule>
 93. The Biomolecular Interaction Network Database. Available at: <http://www.bind.ca/>
 94. Bader GD, Donaldson I, Wolting C, Ouellette BF, Pawson T, Hogue CW. BIND: The Biomolecular Interaction Network Database. *Nucleic Acids Res*. 2001;29:242–245.
 95. Bader GD, Hogue CW. BIND: a data specification for storing and describing biomolecular interactions, molecular complexes and pathways. *Bioinformatics*. 2000;16:465–477.
 96. GenMAPP. Gene Microarray pathway profiler. Available at: <http://www.genmapp.org>
 97. Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR. GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet*. 2002;31:19–20.
 98. National Center for Genome Research. PathDB. Available at: <http://www.ncgr.org/pathdb/>
 99. Fitzhugh R. Thresholds and plateaus in the Hodgkin-Huxley nerve equations. *J Gen Physiol*. 1960;43:867–896.
 100. Noble D. Cardiac action and pacemaker potentials based on the Hodgkin-Huxley equations. *Nature*. 1960;188:495–498.
 101. DiFrancesco D, Noble D. A model of cardiac electrical activity incorporating ionic pumps and concentration changes. *Philos Trans R Soc Lond B Biol Sci*. 1985;307:353–398.
 102. Clancy C, Rudy Y. Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia. *Nature*. 1999;400:566–569.
 103. Greenstein J, Po S, Wu R, Tomaselli G, Winslow RL. Role of the calcium-independent transient outward current I_{to1} in action potential morphology and duration. *Circ Res*. 2000;87:1026–1033.
 104. Jafri S, Rice JJ, Winslow RL. Cardiac Ca^{2+} dynamics: the roles of ryanodine receptor adaptation and sarcoplasmic reticulum load. *Biophys J*. 1998;74:1149–1168.
 105. Puglisi JL, Bers DM. LabHEART: an interactive computer model of rabbit ventricular myocyte ion channels and Ca transport. *Am J Physiol Cell Physiol*. 2001;281:C2049–C2060.
 106. Rice JJ, Jafri MS, Winslow RL. Modeling short-term interval-force relations in cardiac muscle. *Am J Physiol*. 2000;278:H913. Abstract.
 107. Cortassa S, Aon M, Winslow RL, Marban E, O'Rourke B. Modeling mitochondrial Ca^{2+} dynamics and energy metabolism. *Biophys J*. 2002; 82:115a. Abstract.
 108. Cortassa S, Aon M, Winslow RL, Marban E, O'Rourke B. Modeling Ca^{2+} signaling in mitochondria: Ca^{2+} as a mediator of matching between energy supply and demand. *Biophys J*. 2002;82:115a. Abstract.
 109. Cortassa S, Aon M, Marban E, Winslow R, O'Rourke B. An integrated model of cardiac mitochondrial energy metabolism and calcium dynamics. *Biophys J*. 2003;84:1–22.
 110. Winslow R, Greenstein J, Tomaselli G, O'Rourke B. Computational model of the failing myocyte: relating altered gene expression to cellular function. *Phil Trans Roy Soc Lond A*. 2001;359:1187–1200.
 111. Winslow RL, Rice JJ, Jafri MS, Marban E, O'Rourke B. Mechanisms of altered excitation-contraction coupling in canine tachycardia-induced heart failure, II: model studies. *Circ Res*. 1999;84:571–586.
 112. Shaw RM, Rudy Y. Electrophysiologic effects of acute myocardial ischemia: a mechanistic investigation of action potential conduction and conduction failure. *Circ Res*. 1997;80:124–138.
 113. The Luo-Rudy Dynamic (LRd) Model of the Mammalian Ventricular Action Potential. Available at: <http://www.cwru.edu/med/CBRTC/LRdOnline/>
 114. Computational Medicine and Biology Laboratory. Available at: <http://www.cmbi.jhu.edu>
 115. The National Resource for Cell Analysis and Modeling Web site. Available at: <http://www.nrcam.uchc.edu/>
 116. JSIM: A Java-Based Simulation and Animation Environment. Available at: <http://chief.cs.uga.edu/~jam/jsim/>
 117. Computational Library of Cell Models. iCell: interactive cell modeling software. Available at: <http://ssd1.bme.memphis.edu/icell>
 118. The Cellml.org Web site. Available at: <http://www.cellml.org>
 119. Caltech ERATO Kitano Systems Biology Workbench Development Group. The systems biology markup language (SBML). Available at: <http://www.sbml.org/sbml/docs/index.html>