Integrating Bioinformatics and Computational Biology: Perspectives and Possibilities for *In Silico* Network Reconstruction in Molecular Systems Biology

Rui Alves[*,1,2], Ester Vilaprinyo[1,2,3] and Albert Sorribas[1,2]

[1]*Departament Ciències Mèdiques Bàsiques;* [2]*Institut de Recerca Biomèdica de IRBLLEIDA;* [3]*Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Spain*

Abstract: There is a flood of molecular data about many aspects of cellular functioning. This data ranges from sequence and structural data to gene and protein regulation data, including time dependent changes in the concentration. Integration of the different datasets through computational methods is required to extract biological information that is relevant from a systems biology perspective.

In this paper we discuss how different computational tools and methods can be made to work together integrating different types of data, mining these data for biological information, and assisting in pathway reconstruction and biological hypotheses generation. We review the recent body of literature where such integrative approaches are used and discuss automation of data integration and model building to generate testable biological hypotheses. We analyze issues regarding the design of such automated tools and discuss what limitations and pitfalls can be foreseen for the automation and what solutions can computer science and biologists provide to overcome them.

INTRODUCTION

Molecular systems biology is a broad discipline in which computational methods play a central role. Most researchers will agree that molecular systems biology ultimately aims at understanding how molecular systems function when they are assembled. Thus, pathway and circuit reconstruction and mathematical modeling of the corresponding networks, are central issues in this research area. Because of its inherent diversity, different "origins" are acknowledged for the field [1, 2].

From an historical point of view, the use of mathematical models and computer simulations in molecular biology can be traced at least to the early fifties. Its use appears to have been introduced by Britton Chance, Benno Hess, Joe Higgins, David Garfinkel and their colleagues [3-7]. They used modeling to achieve a more systemic understanding of the processes they were studying experimentally. Such processes ranged from the kinetics and mechanism of catalase action to glycolytic oscillations [3-7]. In the sixties and seventies a few other research groups entered the field, and applied engineering control theory to the study of biological systems [8-16]. Also in the sixties, researchers started taking advantage of the computer capabilities for data organization and analysis. The work of Fitsch [17-19], and its further development by Needleman and Wunsch in 1970 [20] and by Smith and Waterman in their seminal 1981 paper [21], would set the stage for the bioinformatics revolution that took place during the last decade in the twentieth century.

This revolution was in all likelihood prompted by an accumulation of gene sequences and protein sequences and structures. The accumulated data required the development of advanced computational tools for their analysis and organization, leading to the bioinformatics burst. As a consequence of the working capacity of the post-genomic high throughput (HTP) techniques and of the capabilities of computer tools, the old reductionism paradigm shifted towards an integrative view of the molecular biology problems. Albeit not new, this integrative view shifts attention from bioinformatics into (molecular) systems biology, which is sensed as a new frontier in biology. This shift was predicted by Bertalanffy and others as early as in 1940 [22]. The current recognition of this new paradigm has been fueled, among others, by Horishi Kitano and John Doyle [22-28].

The size and scope of the accumulated HTP data sets have made it impossible for any one person to analyze and integrate them all, even if one is only interested in a specific molecular biology process on a given organism. However, such integration is fundamental to reconstruct the molecular networks involved in cellular processes and to obtain a systemic perspective of how those networks function. Thus, researchers need tools that assist with processing, filtering, organizing, and appropriately displaying the complex information that is available. Furthermore, automation of analysis and integration is also fundamental for an effective use of that information.

The increasing availability of computational power and storage capabilities, described for example by Moore's law [29], facilitates the creation of CPU-demanding algorithms. These algorithms can be used to build software applications that can manipulate the different available datasets, integrate the information they contain, and display the end result of the analysis in a user friendly format. Web based applications play an important role in making all these software tools available to most scientists. Efficient information mining and integration, and an appropriate display of the results facilitates *in silico* reconstruction of the molecular networks

*Address correspondence to this autor at the Departament Ciències Mèdiques Bàsiques, Universitat de Lleida, Montserrat Roig 2, 25008 Lleida, Spain; E-mail: ralves@cmb.udl.es

1574-8936/08 $55.00+.00

that regulate and execute relevant molecular biology processes.

In silico pathway and circuit reconstruction is an important area within systems biology, and it has consequences for molecular biology, genetics, biotechnology and others. The reconstruction ultimately generates hypotheses regarding the connectivity and dynamic behavior of pathways and circuits that must be validated against experimental predictions data. Different types of *in silico* reconstruction problems exist:

(1) Identifying the pathways, genes and processes/functions that are encoded in the entire genome of an organism. This provides information regarding possible *qualitative* systemic responses of the cell. As a first approach to this reconstruction problem, one may map the annotated genes onto pre-existing charts of metabolism, gene circuits and signal transduction. This action may reveal limitations in adaptive responses of that organism. For example, if no amino acid biosynthetic pathways are found, then this organism is not likely to survive in a medium without and external amino acid source.

(2) Reconstructing the detailed reaction network that exists within specific pathways or circuits (see for example [53-59]). An important aspect in this type of research is the reconstruction of novel pathways and the identification of new components in classical pathways. This is a relevant problem because, either in well characterized organisms or with the sequencing of metagenomes [60], many genes of unknown function exist and several previously unknown pathways are being discovered. This type of reconstruction may allow for more precise predictions regarding how specific parts of the cellular response are regulated and executed, facilitating the creation of mutated organisms with biotechnological interest.

(3) Reconstructing regulatory networks in gene expression and signal transduction. For example, identifying regulatory motifs in DNA will provide indication of what transcription factors may regulate the expression of different proteins and genes that are involved in specific processes (e. g. [61-63]). At the metabolic and signal transduction level, reconstruction of the detailed regulatory networks for the enzymes in a pathway is a requirement for accurate and quantitative prediction of the dynamic cellular behavior in response to environmental challenges.

In this paper we focus on and review the use of HTP datasets for the *in silico* reconstruction of metabolic pathways, signal transduction pathways, and gene circuits.We frame the reconstruction in the context of computational molecular systems biology, an emerging field that merges bioinformatics, and computational biology. Within this context, we shall discuss the data sets and computational approaches that can facilitate that reconstruction. Then, we review the literature for work that uses an integrative computational approach for this reconstruction and discuss automating the integration process. Finally, we discuss the various problems that hinder automation of integration and analysis of the data sets. We discuss in more details some of the technical aspects of the research reviewed in the main paper and provide further references to other sources of information in a supplementary appendix.

WHAT IS OUT THERE? AUTOMATED DATA ANALYSIS IN THE AGE OF OMICS

Pathway reconstruction requires integration of knowledge at many different levels. Before discussing this integration, we briefly characterize each of the various types of datasets that are available for *in silico* reconstruction of molecular biology pathways. Also, we shall discuss the different methods and tools that are more commonly used a) to mine these datasets, and b) to extract network information and facilitate pathway reconstruction.

Bibliomic Data

Bibliographic data has been accumulating for more than a century. Databases such as MEDLINE [64] or the Web of Science Citation Index [65] collect and organize data from this published literature. Through automated keyword analysis, search engines can identify relevant documents in these databases. These documents can be mined for information on genes, pathways, and networks of interest. Most of these documents concern specific, detailed studies about small sets of genes, proteins, RNAs or metabolites within specific organisms. In many cases, this data provides a detailed functional analysis of many individual genes, using methods that are more accurate than those used to obtain HTP data.

Manual and electronic literature analysis has been used for metabolic reconstruction since researchers started creating mathematical models for molecular biology processes [4]. Currently, literature searches are limited at automatically generating a possible network structure for a given molecular process, although they can easily identify papers containing information that is relevant for the reconstruction of that network. In fact, most mathematical models of molecular pathways and circuits are based upon information that was manually retrieved from literature (for example [66] or [67]).

The recent development of tools such as iHOP [68-71] or biobibliometrics [71] allows researchers to automatically reconstruct networks of genes and proteins from automated literature analysis [68-71]. The underlying assumption of these methods is that identifying gene and protein names that co-occur in the same document(s) generates a network of genes/proteins that are functionally related among them. These network reconstruction methods have been used as a starting point to reconstruct pathways such as the Iron Sulfur Cluster (ISC) biogenesis in *Saccharomyces cerevisiae* [55] or to identify genes involved in some types of cancer [72]. However, such a network should be viewed as a low level reconstruction of the molecular pathways that are involved in the processes for which those specific genes are important. Further analysis is advisable before one claims that the automatically generated network is a complete conceptual model of the processes of interest.

The largest fraction of the textual scientific information contained in a paper is unavailable if one mines Medline

exclusively, because the full text of the papers is not included in this database. The PubMed, SCOPUS, PLOS, and BioMedCentral initiatives are crucial in making full text of scientific papers publicly available [73-76]. The ongoing efforts from scientific journal to make their contents fully available on line also contribute to this effort. There is strong awareness that molecular systems biology research will greatly benefit if information mining from text can be automated to a higher degree than it currently is [61, 69, 77].

Sequence Data, Functional Data, and Structural Data

Databases of annotated gene and protein sequences facilitate the subsequent functional annotation of new genomes, through the use of homology comparisons. Gene or protein sequences from different organisms that have very high similarity (homology) are likely to have the same function. Sequence data can also be mined to predict a) regulatory regions and open reading frames of genes, b) RNA genes, and c) targets for regulation by these RNA genes (see supplementary appendix). The accumulated functional knowledge about genes and proteins facilitates the creation of charts for metabolism, signal transduction and gene circuit for the different organisms with fully sequenced and annotated genomes (see for example [78-81]). In such charts, the individual function of a protein is superimposed onto the particular steps where that protein is active. An example where genome annotation has been used to reconstruct the full complement of metabolic pathways for *Lactobacillus plantarum* can be found in [45, 49]. Sequence based annotation is not possible when a new gene is not homologous to any gene of known function. However, if structural information can be gathered for the protein coded by that gene, structural homology comparisons may also facilitate attributing general or specific functions to individual genes, for example using classifications such as SCOP or CATH [82-92]. Knowing the structure of a protein (or RNA gene) can elucidate the mechanism by which these molecules perform their function. Furthermore, having structural templates allows the prediction of structures for other homologous proteins.

Even when homology is not useful for functional annotation, sequence information can still be used to infer some functional information. For example, one can use phylogenetic conservation to investigate possible functions of the genes. The logic behind phylogenetic conservation analysis is as follows. If a set of homologous genes with unknown function is present (absent) with other genes of known function in the same set of genomes, then it is possible that evolution acted simultaneously on that set of genes because somehow they share a function. This may with other genes of known function allow the researcher to predict that some genes are involved in the same processes, although their individual function may remain uncertain. Such an approach has been combined with other lines of evidence, to identify the proteins Yfh1 [93] and Grx5 [58] as being involved in ISC biogenesis. Another example is the application of phylogenetic analysis to the reconstruction of the Coenzyme A biosynthesis pathway in different archaeal genomes [94] and for the reconstruction of parasite nucleotide biosynthesis [95]. A similar method for inferring function is that of finding gene fusion events. Such events imply that two genes share common function [96, 97] and maybe even common regulation, as appears to be the case for example in the biosynthesis of aromatic amino acids in low GC Gram-positive bacteria [98].

Sequence based reconstruction of gene circuits can be more complex than sequence based reconstruction of metabolic and signal transduction pathways. The reason for this is that regulatory sequences in DNA are shorter than coding sequences. This creates a lower signal to noise ratio in their detection, when compared to identifying a full gene. Identification of gene circuits units within a genome can be done for example by searching for regulatory motifs or sequence patterns upstream of gene promoters. Confidence in the accuracy of the regulatory units predicted by this method is increased by finding that such motifs are phylogenetically conserved in different organisms [99-108]. This type of analysis has been used for example to identify novel targets of the Eyeless transcription factor in *Drosophila* [109], to identify transcription factors that regulate human gene expression [110], to identify novel gene circuits for amino acid transport and glucose in *S. cerevisiae* [111, 112] or to predict operons in *Pyrococcus furiosus* [113]. An intense research effort to understand RNA gene circuits is also under way [114-121]. Only when this new layer of regulation is fully understood and integrated into the reconstruction efforts, can one have full reconstruction of gene circuits. Progress in this area is fast and different groups are already reconstructing RNA gene circuits in bacteria and other organisms by integrating the distinct types of information discussed in this section of the review [122-127].

Gene Expression Data

Different types of mathematical analysis allow researchers to infer functional genetic modules and circuits from the analysis of high and low throughput gene expression data [128-136]. Statistical theory and information theory is extensively used to infer regulatory network structures from gene expression data [137-147]. For example, if a gene/protein of unknown function is differentially regulated during some cellular response, then one might infer that this gene is involved in that response. The set of genes responsive to human interferon beta [148, 149] or that of genes involved in the development of rat central nervous system [150] have been reconstituted from microarray data. However, one must always keep in mind that post-translational regulation and fine tuning of enzyme activities may work to modify the importance of changes in gene expression. This may complicate the interpretation of gene expression profiles.

Available gene expression data is often static, in the sense that it is measured at a specific time during the response and no other previous or subsequent measurements are available. Static gene expression data, in general, do not provide significant information about the specific function of a gene in the response. Exceptions might be for example situations where only one protein is missing in a circuit or pathway that is well known. In such cases, if only one gene of unknown function is identified in the expression data, its function is likely to be the one that is missing. Alternatively, it may be possible to infer sequential action of transcription factors in the gene circuit, by using time series data from microarray experiments [151-153]. For example, the network of gene activation in macrophage response [154], the p53

network in human leukemia cell lines [155] and the regulation of galactose biosynthesis [156] have been studied using this type of analysis.

Proteomics Data

Proteomics experiments in which measurements of protein levels and activity are made can also assist in network reconstruction. Having such data is important and complementary to the gene expression data. In fact, proteomic data is required to ultimately refine the network structure of circuits and pathways and assess their functionality *in vivo*. This is so because there are cases in which the levels of the protein that is coded by a gene can change in response to some stimulus, even though expression of that gene is unchanged and no interaction with other proteins has been previously reported. In addition, proteomics analysis may reveal changes in the activity of proteins in responses to challenges that were previously unknown to affect their activity. This kind of information is still lacking for most organisms and cell types.

Proteomics studies are in their infancy, when compared to gene expression studies. Nevertheless, they have been used for pathway reconstruction in various cases. HTP proteomic experiments led to the discovery of networks of direct physical interaction between proteins in different organisms [157-172]. Finding which proteins interact physically with those of unknown function sheds light upon the processes in which the later proteins may be involved in, thus facilitating the reconstruction of their role in the cell. However, one should keep in mind that these experiments may detect false positive interactions and fail to detect real interactions, due to the experiments being made under inappropriate physiological conditions. In addition to detection of protein interactions, proteomic approaches have been used to a) identify the kinases that are necessary for cell cycle progression in *Drosophila* [173], b) reconstruct human phosphorylation networks [174], c) reconstruct the network of growth factor signaling in cancer cells [175], d) study the pheromone response in yeast [176], and e) study signal transduction in plants [177, 178].

It is still unfeasible to use large scale proteomic assays to measure changes in the activity of the full protein complement of an organism. However, the development of protein chips that identify binding of protein to DNA [179] and allow measuring of enzyme activity for whole pathways [180, 181] suggests that such a goal may be attainable in the future.

Metabolomic Data

Metabolomics data can also, in principle, be used to reveal information about pathway and circuit connectivity [182, 183], to estimate parameter values, and to validate and refine models of specific cellular processes [184, 185]. These very important points would justify a wide program of metabolomic experiments. Without this kind of measurements, validation of mathematical models for specific metabolic processes may be difficult. Furthermore, one can infer which reactions and which steps causally precede others in a metabolic or signal transduction network, by measuring the changes in the concentration of metabolites or signaling molecules over time. Such an analysis has been performed to predict the sequence of glycolytic reactions [186] and to further refine the regulatory network of glycolysis in *Lactococcus lactis* [184].

USING MOLECULAR DATA TO PREDICT SYSTEMIC BEHAVIOR

In the previous section we presented a selection of the data that are available for *in silico* inference and reconstruction of network and pathway topology. A given network topology should be able to explain the experimental behavior of a system, if that topology underlies the process that regulates the biological response being measured. However it is often impossible to use common sense for judging how well a given network explains some dynamic response, because the dynamic behavior of biological systems is non-linear. To overcome such a limitation, topological schemas can be used to create mathematical models whose dynamical behavior can be rigorously analyzed and compared to what is experimentally observed. This validation process is fundamental in testing the correctness of *in silico* pathway reconstruction.

Building Mathematical Models

Mathematical models can be created for networks of different scales. Some researchers are interested in modeling the entire metabolic network found in a genome (e.g. [31-52]. Such models have been analyzed to predict both, growth characteristics, and essentiality of genes in different organisms using flux balance analysis (FBA) [35, 38, 39, 46]. Although in mathematical terms FBA models are linear, the qualitative predictions of phenotype have so far held for between 60% and 80% of the genes [39, 46, 187]. These models inappropriately represent dynamic regulatory effects, which could account for some of the erroneous predictions (see [188] for details).

Smaller scale models that accurately account for dynamic regulation are important for testing more detailed hypotheses about the functioning of specific processes. For example, the regulatory structure of the Iron-Sulfur Cluster biogenesis pathway in *S. cerevisiae* has been validated by comparing the dynamic behavior of mathematical models for alternative regulatory topologies to experimental results [55]. Similarly, models of cell cycle have been validated by comparing its dynamic behavior to experimental results [189-191], and models of the pentose phosphate pathway have been validated by predicting the best targets for treatment of metabolic diseases [192-195].

Statistical methods can be used how well alternative networks can reproduce experimental behavior, if sufficient information is available. For example, one can use optimization algorithms to determine which network more accurately fits the known quantitative behavior of a system [196-198]. Qualitative statistical methods should be used for this evaluation, if the available experimental data is not quantitative. For example tree decision algorithms have been used to analyze the qualitative behavior of signal transduction modeling [199, 200].

Different types of mathematical modeling can be used to study the same biological problem. The choice of a particular type depends on the level of complexity and details one wishes to consider, on the available data, on personal prefer-

ence, etc. Differential equations, either ordinary (for well stirred systems) or partial (for systems with spatial differentiation), are used to define mathematical models that can quantitatively simulate the dynamics of a molecular system [4, 201]. This type of mathematical approach is still the most prevalent choice as a modeling tool, although stochastic approaches are becoming popular. Alternative approaches are further discussed in the supplementary appendix.

Bottlenecks in Model Building

One limitation to the creation of mathematical models is that, for many processes, the parameter values are unknown. Even when parameter values are available, locating them and evaluating if the experimental conditions under which they have been measured are appropriate for our modeling purposes is not easy. Furthermore, automatically identifying documents from the literature in which such information is available is still a task that is hard to automate. Another limitation in creating a mathematical model is when the knowledge about the actual mechanism that underlies many biological processes is lacking. This makes it impossible to derive classical Michaelis-Menten like expressions for the kinetics of those processes. Mathematical representations based on approximate formalisms, such as power-law models, help side-step this lack of information (see supplementary appendix and [202]). Creating a mathematical model is a very work intensive task that is prone to human error. As the dimension of a model increases, automated model building becomes important in decreasing modeling errors. General automated model building is possible only by using structured and systematic formalisms.

Most of these formalisms are mathematical approximations that simplify the exact kinetic expressions representing the dynamical behavior of individual molecular biology processes and reduce the dimension of the system of equations to be solved (for example [203]). Using structured formalisms that are based in approximation theory also sidesteps the lack of knowledge regarding mechanism while building a mathematical model [204]. These approximate representations are accurate over a varying range of values about their operating point [205] and they can be invaluable as canonical formalisms that facilitate automated generation, analysis, and exchange of mathematical models [204]. For a review of these ideas see [202].

INTEGRATION OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY FOR AUTOMATED NETWORK EXTRACTION AND MODEL BUILDING

Automatic model building that is based on the topological representation of a network requires integration of information at different levels. Fig. **1** outlines in a flow chart how one can integrate the bioinformatics data mining process to the model building process in order to complete the *in silico* reconstruction of networks. In general, the integration needs to be highly parallelized, allowing for flexible exclusion of one or more types of datasets and for an appropriate interaction with the user.

Outline of the Integration

A first layer of information (sequence data, structural data, literature data, gene expression data, proteomics data and metabolomics data) available in HTP data sets and published material can be integrated to derive a second layer of processed data. This second data layer contains information about a) gene/protein function and interactions, b) pathway and gene circuit topology, and c) parameter values. The second layer of data also receives direct experimental inputs at different levels. For example, experimentally determined parameter values can be directly integrated into this layer. The data from the second layer can in turn be integrated to generate alternative topological schemas that describe how the pathways and circuits of interest may function. Statistical algorithms, such as decision tree methods or Bayesian analysis, can be used at different stages of the reconstruction to rank the relevance of the information and give users an objective means of deciding what is the most appropriate information to include in the reconstruction.

The topological schemas can be used either to refine preexisting mathematical models or to generate new ones. The dynamical behavior of alternative models can then be compared to the experimental behavior of the system. By using statistical methods to analyze these comparisons, one may objectively rank which schema are more likely to correctly describe the topology that underlies the observed behavior. (Fig. **2**) details how the analysis and integration of the different datasets can be accomplished and used for pathway and circuit reconstruction.

The Role of Text-Mining

Fig. **2A** suggests how literature analysis (***bibliomics***) can be used to generate conceptual models for pathway and circuit reconstruction. Web crawling robots, using Artificial Intelligence (AI) algorithms, can perform automated keyword and contextual analysis of information contained either in literature databases or in live web documents. They can also identify lists of genes that appear to be relevant for the process of interest and generate a tentative topological network by suggesting functions for those genes.

Automated and contextual text analysis may be extremely important to locate information about parameter values and enzyme kinetics and in automatically integrating such information into models. However, human intervention is still the most accurate way for mining this type of information from the literature, which suggests that the need for human curation should always be kept in mind during the different stages of the reconstruction process.

If appropriate standards of model and parameter reporting are defined, automated text mining of parameter values from the literature may be possible. One must be cautious regarding these standards. Many may be tempted to use parameter values and models that are automatically found, without considering the appropriateness of the models for the goals one has and the compatibility between different parameter determinations.

The Role of Mining "omics"

Fig. **2B** suggests how functional information, sequence information and gene expression information (***genomics***) can be used within a global strategy for pathway reconstruction. By processing sequence information one can automatically identify genes and proteins that participate in a given process

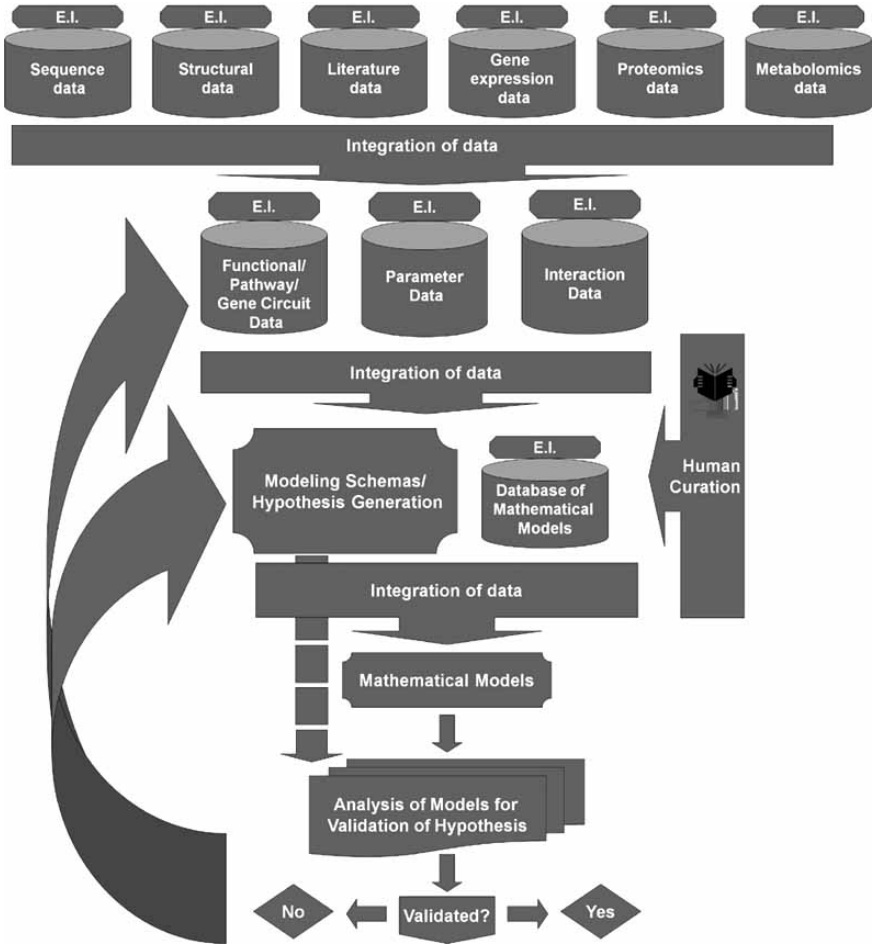


Fig. (1). Integrating the different datasets for pathway reconstruction: An overview. E.I. stands for direct experimental input to a given class of data. Different types of primary data can be used and integrated to infer functional information. Functional information can also be experimentally generated. All the relevant information can be integrated, thus generating conceptual schemas that can be used to generate and test hypotheses regarding the behavior of the molecular system of interest.See text for details.

and generate functional information for those molecules. In parallel, gene expression data can, in principle, also be used to derive some form of causal relationship between changes in the expression of the different genes. Besides identifying/confirming genes involved in the processes of interest, the information processing shown in Fig. **2B** can also be used to generate alternative architectures for the topology of gene circuits and pathways.

Fig. **2C** suggests that ***proteomics*** and ***interaction data*** can be used to derive a network of protein and gene interactions. In addition to identifying or confirming previously identified proteins, the interaction data can provide additional hypotheses regarding the previously unknown connectivity in a pathway.

Proteomics and metabolomics data can also be used to infer the sequence in which different proteins act in a pathway and how the activity of the different steps is regulated. "Omics" data may also contain essential information for identifying appropriate parameters for the mathematical models [2], as long as structured formalisms are used to write the model equations and sufficiently dense time series are available in the data. For more information on this subject see the supplementary appendix and references therein.

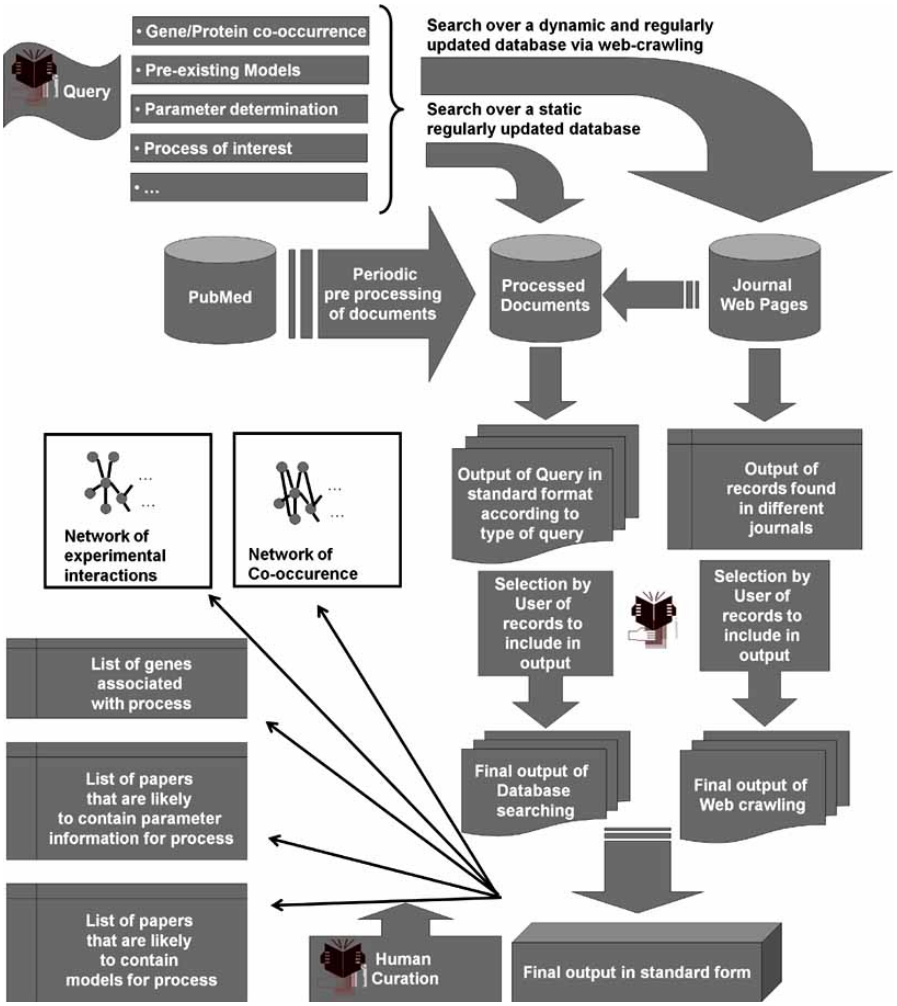
From Bioinformatics to Computational Systems Biology Through Model Building

Fig. **2D** details the final stage of the integration between bioinformatics and computational systems biology. The user can attribute alternative roles in the network to genes even if they have not been automatically assigned to specific reaction steps or regulatory interactions. Once a final set of alternative schemas for the pathways and circuits of interest are created, the users must decide upon the level of detail that is necessary for their research. Automated search of databases with pre-existing models can find previously build mathematical models of the reconstructed network, if they exist. Once the models are created, they can then be analyzed, validated, and refined. The process of model validation, described in previous sections, can stimulate new research, assisting in the design of experiments that clarify unsolved questions and allowing for rational testing of hypotheses about the systemic behavior of the network.

Examples of Bioinformatics and Computational Biology Integration in Molecular Biology

It is consensual that the use and integration of different datasets, combined with mathematical modeling is important

A


Query
• Gene/Protein co-occurrence
• Pre-existing Models
• Parameter determination
• Process of interest
• ...
Search over a dynamic and regularly updated database via web-crawling
Search over a static regularly updated database
PubMed
Periodic pre processing of documents
Processed Documents
Journal Web Pages
Network of experimental interactions
Network of Co-occurence
Output of Query in standard format according to type of query
Output of records found in different journals
Selection by User of records to include in output
Selection by User of records to include in output
List of genes associated with process
List of papers that are likely to contain parameter information for process
List of papers that are likely to contain models for process
Final output of Database searching
Final output of Web crawling
Human Curation
Final output in standard form

B

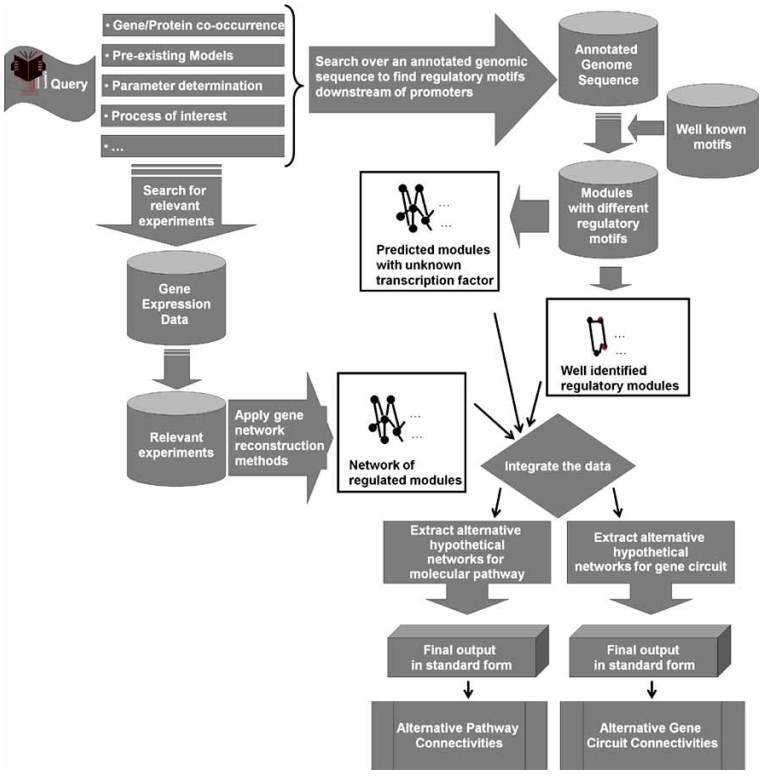
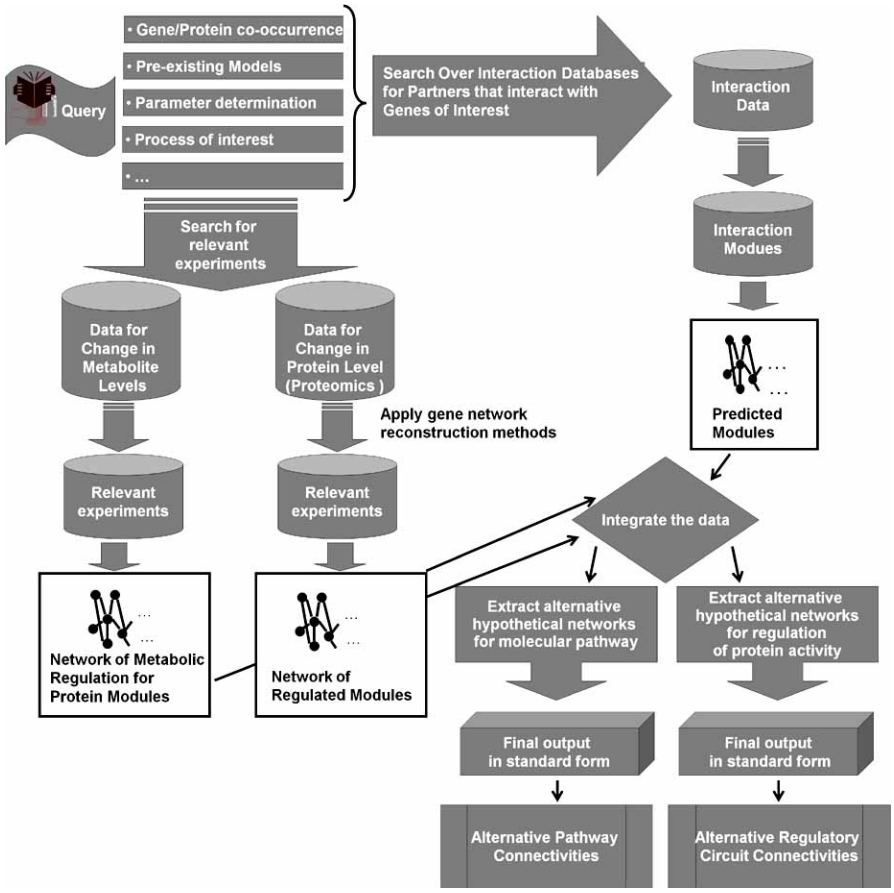

Query
• Gene/Protein co-occurrence
• Pre-existing Models
• Parameter determination
• Process of interest
• ...
Search over an annotated genomic sequence to find regulatory motifs downstream of promoters
Annotated Genome Sequence
Well known motifs
Search for relevant experiments
Predicted modules with unknown transcription factor
Modules with different regulatory motifs
Gene Expression Data
Well identified regulatory modules
Relevant experiments
Apply gene network reconstruction methods
Network of regulated modules
Integrate the data
Extract alternative hypothetical networks for molecular pathway
Extract alternative hypothetical networks for gene circuit
Final output in standard form
Final output in standard form
Alternative Pathway Connectivities
Alternative Gene Circuit Connectivities

Fig. (2). Contd…

C



D

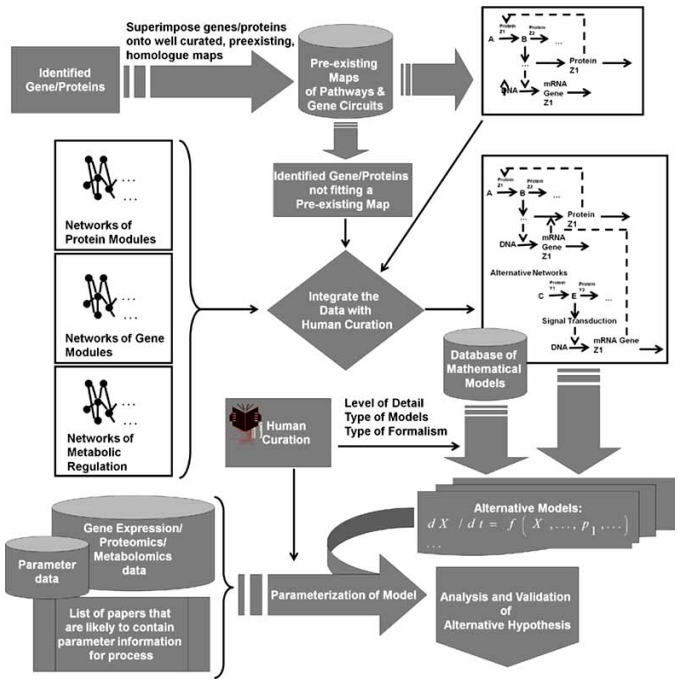


Fig. (2). Detailed schemes for the data mining of the different datasets for pathway and circuit reconstruction. **A –** Mining the literature. Automated analysis of abstracts and full text of papers can assist in reconstructing functional molecular networks of the cell. **B –** Mining gene expression data and genome sequence. Combining the analysis of gene expression data with that of sequence data, functional networks of gene regulation can be generated. **C –** Mining Proteomics and Metabolomics datasets. Proteomics and metabolomics information can be combined to derive protein modules with causal interactions. **D –** Integrating the data to generate hypotheses and validate models. Integrating the data from **A**, **B**, and **C** one can generate mathematical models that can be used to generate and test hypotheses regarding the behavior of molecular systems. See text for a more detailed discussion

in facilitating and improving the accuracy of *in silico* reconstruction of molecular pathways in cells [37, 41, 53-55, 61, 206-208]. For example, Alves and Sorribas have combined the use of literature analysis, phylogenetic profiling, analysis of sequences, and protein interaction prediction to reconstruct the regulatory network of mitochondrial ISC biogenesis in yeast [53-55, 58]. Bas Teusink's group has been developing a method, AUTOGRAPH, that combines sequence homology analysis with the existence of well curated metabolic maps to reconstruct the complete metabolic networks of new genomes. They have applied their method to the reconstruction of *L. lactis* metabolism [41]. Su *et al.* [209, 210] reconstruct both the pathway of phosphate assimilation and the gene circuits that regulate the expression of that pathway in *Synechococcus*, by combining genomic information with information about interactions between different genes and proteins. A combination of literature analysis and microarray data analysis has also been used to derive a regulatory network for *E. coli* and to test the consistency of microarray data based predictions [211]. There is an overlap of approximately 80% between the networks derived by the two methods. A reconstruction analysis of the regulatory network for the galactose biosynthesis pathway in yeast successfully predicts the genes with stronger regulatory influence on the pathway. The analysis has also been made by combining microarray data and protein interaction data [156]. A combination of time series analysis of gene expression and *in silico* prediction of transcription factor biding sites has been used to define regulatory modules in the inflammatory response of the macrophage, suggesting novel roles for the transcription factors ATF3 and NRF2 [154]. MAPK signaling pathways in the human blood fluke have been reconstructed by integrating phylogenetic conservation analysis and experimental gene expression measurements [212]. An analysis of the human phosphoproteome by combining consensus substrate motifs with context modeling was used for improved prediction of cellular kinase-substrate relationships [174, 213, 214].

Currently, the process of combining different datasets to generate testable biological hypotheses lacks a well defined structure and can only be partially automated. Nevertheless, there is constant progress along the road towards such automation, and for example, both Su *et al.* [209, 210] and Alves and Sorribas [55] propose and apply structured integrative approach for network reconstruction in molecular biology. An area where integration of bioinformatics and computational biology may be a powerful tool is that of synthetic biology. We will not discuss this subject any further in this paper and refer to the literature for reviews on the subject [215-221].

CHALLENGES TO AUTOMATION IN COMPUTATIONAL SYSTEMS BIOLOGY

Automatically integrating the information from the different datasets for pathway and circuit reconstruction in molecular systems biology is not trivial. The design of a global solution for this integration requires careful consideration of the goals, challenges, and limitations of the available a) data, b) data mining methods, and c) mathematical models.

At a first glance, one might think that the amount of data being generated by HTP methods is the most difficult challenge for the integration process. However, enough computational power is available to deal with this problem and data accumulation may not be a major issue. In our perspective, the major challenges are likely to be a) the definition of the necessary information content for a given type of data, b) the development of universal standards for the reporting and deposition of that data into databases, and c) the integration of information mined from different types of data into a coherent whole. It is crucial that the relevant information that is needed to address a given biological problem is well organized and easily accessible. Developing and applying standards in data reporting plays a central role in facilitating the automation of integrative approaches, because the existent of regularities in data structure is fundamental for the development of efficient computational methods.

Minimal Information Content and Data Reporting Standards

Achieving consistency in data organization, classification and storage into databases is crucial for achieving automated data integration. An important step in achieving such consistency is deciding what information to store in databases and how to report and organize that information.

It is not easy to define the minimal information required to describe the validity and scope of the deposited data, because the relevant information will change depending on the problem one wants to address when using the deposited such data. In some cases the data will have less than the minimal information required to correctly assess the conditions in which the experiment was done and the caveats that apply to the results. In other cases there will exist an excess of information that may be confusing and difficult to organize for one's purpose. An additional problem in comparing similar data from different origins is the following. When different groups replicate experiments using similar techniques, most of the times there are differences either in the exact conditions of the experiment or in the method for data acquisition and treatment. This may lead to contradictory results for what is apparently the same experiment.

If one wants to potentiate automated data mining and integration, a uniform language and ontology that can be used to report all experiments and data of a given type is needed. This ontology and language should contain sufficient information to decide if two data sets replicate the same experiment or not. Whatever standard ontologies evolve for the different datasets, such classification schemes must a) be self consistent, b) be applicable to any organism, c) have the fullest possible coverage of biological/experimental functions, d) adapt to new knowledge and information, and e) be easily integrated with ontologies for other types of datasets.

Several ontologies have been developed to describe the different datasets described in the previous sections (for example [222-233]). No set of ontologies is yet universally accepted. However, the Open Biomedical Ontologies (OBO), and the Gene Ontology (GO) are becoming *de facto* standards in the area [227]. Creating an ontology or a report language is not an easy task and the lack of an appropriate initial structure may compromise future utility. For example,

the GO has an uneven amount of information given for different classes of terms at the same depth of classification [234], which may become a problem in future. An area where information content standards are essential if one wants to facilitate automated model building is that of reporting parameter values for kinetic and thermodynamic experiments (see above section 3 and supplementary appendix). Standard report languages that, to some extent integrate relevant ontology information, have been proposed for HTP experiments [235], mathematical models [24, 236], gene that is expression data [237], mass spectrometry data [http://psi-dev.sourceforge.net/ms/xml/mzdata/mzdata.html], protein interaction data [160], and metabolomics data [225]. It is unclear if any of these languages obeys appropriate criteria for minimum information content and they are still far from being universally applied. However, computer-based algorithms and methods will only be able to automatically cross-correlate all available information in an efficient and error-free way when reporting of HTP experiments and data has converged to some standard.

Imagine that such standards exist and that one wants to automatically build the topology of the complete network of molecular interactions for an organism. An appropriate classification of the different data would, in principle: a) organize all enzymes into metabolic pathways; b) organize all signal transduction proteins, into signal transduction pathways; and c) organize all transcription factors (TF) and genes with binding sites for those TF, into gene circuits (Fig. **3**). Whenever a new genome is sequenced and annotated, its genes can be inserted into the maps, displaying which pathways and circuits are present in that specific organism. As differences in regulation between organisms are found or new pathways and circuits are discovered and reported, the classifications and maps themselves can be updated.

Informational Models for Integrated Data Analysis

There are different organizational models that can be used to store and integrate molecular biology data. One model is that of a central warehouse that stores database information (e.g. [238-243]). These databases are then linked to a server that provides different types of analytical tools. These tools allow users to mine and, up to a point, integrate information from the databases. Often, these central warehouses have a few mirrors distributed over the world wide web. Another type of integration service that is often found on the internet is the Metaserver. These servers, which usually analyze one type of data (e.g. [244-266]), submit queries to many different central warehouses and then process and integrate all the outputs.

Another organizational model for information storage and analysis is that of distributed sources of information and analytical services. Such an approach underlies GRID technology. At this moment there are at least three GRID-based approaches to the development of integrative platforms: BioMOBY [267-271], myGRID [272] or caBIO [273]. They rely on different levels of analysis and aim at integrating services and databases provided by different providers on the web. Such distributed architectures take advantage of the decentralized nature of the information and crawl over the internet space, identifying and accessing different types of data. The data are then pipeline into appropriate services, also decentralized and identified over the internet or in a central repository of services, to perform the analysis required by the user. However, there is still a long way to go before full automated integration of all available types of biological data that are available on the web is achieved.

Currently, the semantic web or web 3.0 is under development [274]. This new technology is expected to integrate information using machine learning methods. The effect of such technology in facilitating automated analysis and model building in molecular biology is potentially large [274-281].

Constraints to Automated Reconstruction of Biological Circuits

It is a truism that no one can be an expert in everything, which makes collaborative work fundamental in an integrative endeavor such as pathway reconstruction in molecular systems biology. However, it is also true that sometimes the demand for collaborative work placed upon researchers by both funding agencies and lack of time to familiarize themselves with different areas, rather than enabling research synergisms, is a hindrance to the progress of the research. Development of a structured integrative approach for network reconstruction and hypotheses evaluation would be an important contribution towards increasing the synergism among groups interested in different parts of the same biological problem.

Independently of the informational model(s) for data reporting, storage and integration, the user interface must be designed taking into account the target audience(s) for the different services. That audience will potentially be as diverse as the entire molecular biology community itself is. Hence, an integrative software service must take into account users with different expertise and different goals. Therefore, a flexible pipeline should exist for data processing and analysis. This pipeline should allow for different types of reconstruction questions to be asked within the integrative architecture. Some users will be satisfied by clicking in buttons and obtaining the final result, while other will want to manipulate the parameters and algorithms behind the interface for their own purposes and for refining standard analysis. There is a need to provide tools that are sufficiently constrained to reduce scientifically unsound use and sufficiently flexible to allow expert users to refine methods and data inputs. It is possible that the development of different tools for the different audiences is the most adequate way to address this problem.

Furthermore, one must be prepared to allow the users to perform different levels of reconstruction and to introduce new qualitative and quantitative information based on their own expertise. While some will be interested in the static reconstruction of pathways and circuits, others will be interested in the dynamic aspects of that reconstruction. Additionally, while some will be interested in a broad picture of what an organism can do, others will want to pursue a more detailed analysis of parts of the cellular response of that organism and only be interested in AI specific pathways. Thus, if one wishes to develop a methodology and implement it in a set of tools that can be useful to a wide audience and used in a scientifically sound way, one has to take several, sometimes contradictory, constraints into account. Human cura-

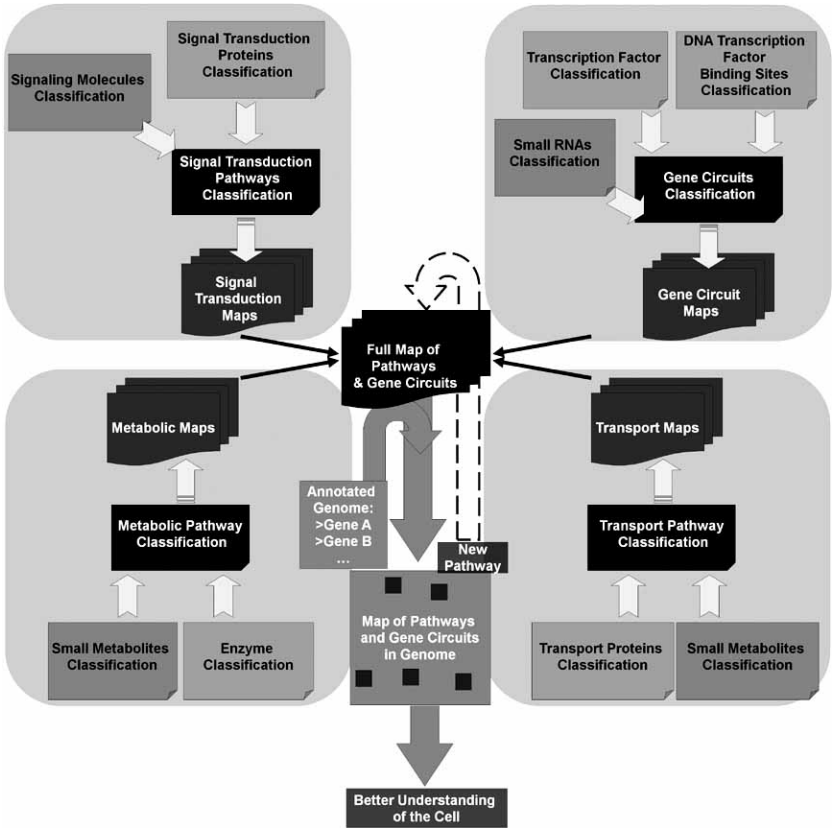


Fig. (3). Creating and integrating maps for pathways and gene circuits and applying them to the reconstruction of pathways in a genome will enable a better understanding of cellular physiology and behavior. Pre-existing gene circuits, metabolic pathways, transport pathways and signal transduction pathways can be used as a template for the assembly of the cellular pathways in newly annotated genomes. New pathways that are found in those genomes can be added to the set of pre-existing pathways. See text for a more detailed discussion

tion is and will continue to be an important part in the process, at even when AI development is able to generate tools that, to some degree, can replace expert intervention.

From a technical point of view, it is unlikely that a single research group will be able to come with an all-encompassing software platform that will allow for such an enormous diversity of reconstructions and goals. Thus, one of the requirements for such platforms is that they are open to external contributions. However, this openness must be somewhat more controlled than what is one finds in most of the open source projects that exist nowadays. Contributing programmers would have to adhere to some programming, organizational, and nomenclature standards, in a structure similar to that of wikipedia. The reason for such constraints is that in many cases there exists both duplication of efforts and inconsistent nomenclature even in complementary packages. An example for both these problems are many packages for the analysis of HTP data for developed the R platform.

Finally, an integrative platform should include some form of machine learning algorithms to analyze older datasets. It is unlikely that many resources are dedicated to reformatting data from previous decades, making that data compatible with whatever standards percolate for each type of dataset. Trainable machine learning algorithms are currently the best hope of mining the older data efficiently. This is an important challenge for the AI community.

SUMMARY AND CONCLUSIONS

In this work we analyze the *in silico* reconstruction of cellular pathways within the context of molecular systems biology. We start with a discussion of the data sets that are available for this reconstruction in the post genomic era. We describe how systematic integration of the different types of information can be used to create mathematical models that generate testable biological hypotheses about the functioning of the reconstructed system. We then review examples of how such data have been integrated to reconstruct pathways and obtain a systemic understanding of how those pathways work. Finally, we provide a general discussion of how different bioinformatics and computational biology tools can be integrated for automated model generation. We are aware that a full automation of this process may not be possible. However, if we are ever to approach such a goal, at least two issues must be addressed:

a) Better and uniform standards of reporting for the different datasets.

b) Better machine learning methods/artificial intelligence algorithms to facilitate automation of the analysis.

Any improvement in these areas will translate into a more efficient and more accurate network reconstruction and facilitate the systemic understanding of molecular pathways and circuits.

ACKNOWLEDGEMENTS

This work has been partially supported by grants BFU2005-0234BMC and BFU2007-62772/BMC of the Spanish Ministerio de Educación y Ciencia. RA was supported by a Ramon y Cajal award from the Spanish Ministerio de Educacin y Ciencia. We also thank two anonymous reviewers for helpful suggestions that significantly improved this paper.

REFERENCES

[1] Eisenberg D, Marcotte E, McLachlan AD, Pellegrini M. Bioinformatic challenges for the next decade(s). *Philos Trans R Soc Lond B Biol Sci* **2006**; 361: 525-527.

[2] Goel G, Chou IC, Voit EO. Biological systems modeling and analysis: a biomolecular technique of the twenty-first century. *J Biomol Tech* **2006**; 17: 252-269.

[3] Chance B, Garfinkel D, Higgins J, Hess B. Metabolic control mechanisms. 5. A solution for the equations representing interaction between glycolysis and respiration in ascites tumor cells. *J Biol Chem* **1960**; 235: 2426-2439.

[4] Chance B, Greenstein DS, Higgins J, Yang CC. The mechanism of catalase action. II. Electric analog computer studies. *Arch Biochem* **1952**; 37: 322-339.

[5] Garfinkel D, Garfinkel L, Pring M, Green SB, Chance B. Computer applications to biochemical kinetics. *Annu Rev Biochem* **1970**; 39: 473-498.

[6] Garfinkel D, Hess B. Metabolic Control Mechanisms. Vii.a Detailed Computer Model of the Glycolytic Pathway in Ascites Cells. *J Biol Chem* **1964**; 239: 971-983.

[7] Hess B, Boiteux A. Mechanism of glycolytic oscillation in yeast. I. Aerobic and anaerobic growth conditions for obtaining glycolytic oscillation. *Hoppe Seylers Z Physiol Chem* **1968**; 349: 1567-1574.

[8] Heinrich R, Rapoport TA. Linear theory of enzymatic chains; its application for the analysis of the crossover theorem and of the glycolysis of human erythrocytes. *Acta Biol Med Ger* **1973**; 31: 479-494.

[9] Heinrich R, Rapoport TA. A linear steady-state treatment of enzymatic chains. Critique of the crossover theorem and a general procedure to identify interaction sites with an effector. *Eur J Biochem* **1974**; 42: 97-105.

[10] Heinrich R, Rapoport TA. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur J Biochem* **1974**; 42: 89-95.

[11] Kacser H. Kinetic models of development and heredity. *Symp Soc Exp Biol* **1960**; 14: 13-27.

[12] Kacser H, Burns JA. The control of flux. *Symp Soc Exp Biol* **1973**; 27: 65-104.

[13] Rapoport TA, Heinrich R, Jacobasch G, Rapoport S. A linear steady-state treatment of enzymatic chains. A mathematical model of glycolysis of human erythrocytes. *Eur J Biochem* **1974**; 42: 107-120.

[14] Savageau MA. Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation. *J Theor Biol* **1969**; 25: 370-379.

[15] Savageau MA. Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J Theor Biol* **1969**; 25: 365-369.

[16] Savageau MA. Biochemical systems analysis. 3. Dynamic solutions using a power-law approximation. *J Theor Biol* **1970**; 26: 215-226.

[17] Fitch WM. An improved method of testing for evolutionary homology. *J Mol Biol* **1966**; 16: 9-16.

[18] Fitch WM. Locating gaps in amino acid sequences to optimize the homology between two proteins. *Biochem Genet* **1969**; 3: 99-108.

[19] Fitch WM. Further improvements in the method of testing for evolutionary homology among proteins. *J Mol Biol* **1970**; 49: 1-14.

[20] Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **1970**; 48: 443-453.

[21] Smith TF, Waterman MS, Fitch WM. Comparative biosequence metrics. *J Mol Evol* 1981; 18: 38-46.

[22] Bertalanffy K. Der Organismus als physikalisches System betrachtet**.** *Die Naturwissenschaften* **1940**; 28.

[23] Csete ME, Doyle JC. Reverse engineering of biological complexity. *Science* **2002**; 295: 1664-1669.

[24] Hucka M, Finney A, Bornstein BJ, *et al.* Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *Syst Biol (Stevenage)* **2004**; 1: 41-53.

[25] Kitano H. Systems biology: a brief overview. *Science* **2002**; 295: 1662-1664.

[26] Kitano H, Oda K, Kimura T, Matsuoka Y, Csete M, Doyle J, Muramatsu M. Metabolic syndrome and robustness tradeoffs. *Diabetes* **2004**; 53 Suppl 3: S6-S15.

[27] Stelling J, Sauer U, Szallasi Z, Doyle FJ, 3rd, Doyle J. Robustness of cellular functions. *Cell* **2004**; 118: 675-685.

[28] Tanaka R, Csete M, Doyle J. Highly optimised global organisation of metabolic networks. *Syst Biol (Stevenage)* **2005**; 152: 179-184.

[29] Mollick E. Establishing Moore's law. *Ieee Ann Hist Comput* **2006**; 28: 62-75.

[30] Proteomics. *Nat Biotechnol,*: IT45-46.

[31] Allen TE, Price ND, Joyce AR, Palsson BO. Long-range periodic patterns in microbial genomes indicate significant multi-scale chromosomal organization. *PLoS Comput Biol* **2006**; 2: e2.

[32] Duarte NC, Becker SA, Jamshidi N, *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA* **2007**; 104: 1777-1782.

[33] Duarte NC, Herrgard MJ, Palsson BO. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* **2004**; 14: 1298-1309.

[34] Duarte NC, Palsson BO, Fu P. Integrated analysis of metabolic phenotypes in Saccharomyces cerevisiae. *BMC Genomics* **2004**; 5: 63.

[35] Edwards JS, Ibarra RU, Palsson BO. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat Biotechnol* **2001**; 19: 125-130.

[36] Feist AM, Scholten JC, Palsson BO, Brockman FJ, Ideker T. Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. *Mol Syst Biol* **2006**; 2: 2006 0004.

[37] Francke C, Siezen RJ, Teusink B. Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol* **2005**; 13: 550-558.

[38] Ibarra RU, Edwards JS, Palsson BO. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **2002**; 420: 186-189.

[39] Ibarra RU, Fu P, Palsson BO, DiTonno JR, Edwards JS. Quantitative analysis of Escherichia coli metabolic phenotypes within the context of phenotypic phase planes. *J Mol Microbiol Biotechnol* **2003**; 6: 101-108.

[40] Mahadevan R, Bond DR, Butler JE, *et al.* Characterization of metabolism in the Fe(III)-reducing organism Geobacter sulfurreducens by constraint-based modeling. *Appl Environ Microbiol* **2006**; 72: 1558-1568.

[41] Notebaart RA, van Enckevort FH, Francke C, Siezen RJ, Teusink B. Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics* **2006**; 7: 296.

[42] Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R. Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *J Biol Chem* **2007**; 282: 28791-28799.

[43] Price ND, Thiele I, Palsson BO. Candidate states of Helicobacter pylori's genome-scale metabolic network upon application of "loop law" thermodynamic constraints. *Biophys J* **2006**; 90: 3919-3928.

[44] Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR**).** *Genome Biol* **2003**; 4: R54.

[45] Rossell S, van der Weijden CC, Lindenbergh A, *et al.* Unraveling the complexity of flux regulation: a new method demonstrated for nutrient starvation in Saccharomyces cerevisiae. *Proc Natl Acad Sci USA* **2006**; 103: 2166-2171.

[46] Schilling CH, Covert MW, Famili I, *et al.* Genome-scale metabolic model of Helicobacter pylori 26695. *J Bacteriol* **2002**; 184: 4582-4593.

[47] Schilling CH, Palsson BO. Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis. *J Theor Biol* **2000**; 203: 249-283.

[48] Teusink B, van Enckevort FH, Francke C, *et al.* In silico reconstruction of the metabolic pathways of Lactobacillus plantarum:

comparing predictions of nutrient requirements with those from growth experiments. *Appl Environ Microbiol* **2005**; 71: 7253-7262.
[49] Teusink B, Wiersma A, Molenaar D, *et al.* Analysis of growth of Lactobacillus plantarum WCFS1 on a complex medium using a genome-scale metabolic model. *J Biol Chem* **2006**; 281: 40041-40048.
[50] Thiele I, Price ND, Vo TD, Palsson BO. Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet. *J Biol Chem* **2005**; 280: 11683-11695.
[51] Thiele I, Vo TD, Price ND, Palsson BO. Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. *J Bacteriol* **2005**; 187: 5818-5830.
[52] Yeung M, Thiele I, Palsson BO. Estimation of the number of extreme pathways for metabolic networks. *BMC Bioinformatics* **2007**; 8: 363.
[53] Alves R, Herrero E, Sorribas A. Predictive reconstruction of the mitochondrial iron-sulfur cluster assembly metabolism. II. Role of glutaredoxin Grx5. *Proteins* **2004**; 57: 481-492.
[54] Alves R, Herrero E, Sorribas A. Predictive reconstruction of the mitochondrial iron-sulfur cluster assembly metabolism: I. The role of the protein pair ferredoxin-ferredoxin reductase (Yah1-Arh1**)**. *Proteins* **2004**; 56: 354-366.
[55] Alves R, Sorribas A. In silico pathway reconstruction: Iron-sulfur cluster biogenesis in Saccharomyces cerevisiae. *BMC Syst Biol* **2007**; 1: 10.
[56] Haugen AC, Kelley R, Collins JB, *et al.* Integrating phenotypic and expression profiles to map arsenic-response networks. *Genome Biol* **2004**; 5: R95.
[57] Panina EM, Vitreschak AG, Mironov AA, Gelfand MS. Regulation of aromatic amino acid biosynthesis in gamma-proteobacteria. *J Mol Microbiol Biotechnol* **2001**; 3: 529-543.
[58] Vilella F, Alves R, Rodriguez-Manzaneque MT, *et al.* Evolution and cellular function of monothiol glutaredoxins: involvement in iron-sulphur cluster assembly. *Comparative and Functional Genomics* **2004**; 5: 328-341.
[59] Workman CT, Mak HC, McCuine S, *et al.* A systems approach to mapping DNA damage response pathways. *Science* **2006**; 312: 1054-1059.
[60] Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P. Prediction of effective genome size in metagenomic samples. *Genome Biol* **2007**; 8: R10.
[61] Herrgard MJ, Covert MW, Palsson BO. Reconstruction of microbial transcriptional regulatory networks. *Curr Opin Biotechnol* **2004**; 15: 70-77.
[62] Herrgard MJ, Lee BS, Portnoy V, Palsson BO. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome Res* **2006**; 16: 627-635.
[63] Herrgard MJ, Palsson BO. Untangling the web of functional and physical interactions in yeast. *J Biol* **2005**; 4: 5.
[64] Richards DE. Medline at thirty-five. *Evid Based Dent* **2006**; 7: 89.
[65] Lowe J, Chapman S. The citation index: what it is and what it isn't. *Aust N Z J Public Health* **2000**; 24: 5.
[66] Heinemann M, Kummel A, Ruinatscha R, Panke S. In silico genome-scale reconstruction and validation of the Staphylococcus aureus metabolic network. *Biotechnol Bioeng* **2005**; 92: 850-864.
[67] Li Z, Chan C. Integrating gene expression and metabolic profiles. *J Biol Chem* **2004**; 279: 27124-27137.
[68] Fernandez JM, Hoffmann R, Valencia A. iHOP web services. *Nucleic Acids Res* **2007**; 35: W21-26.
[69] Hoffmann R, Krallinger M, Andres E, Tamames J, Blaschke C, Valencia A. Text mining for metabolic pathways, signaling cascades, and protein networks. *Sci STKE* **2005**, 2005: pe21.
[70] Hoffmann R, Valencia A. Implementing the iHOP concept for navigation of biomedical literature. *Bioinformatics* **2005**; 21 Suppl 2: ii252-ii258.
[71] Stapley BJ, Benoit G. Biobibliometrics: information retrieval and visualization from co-occurrences of gene names in Medline abstracts. *Pac Symp Biocomput* **2000**; 529-540.
[72] Higgins ME, Claremont M, Major JE, Sander C, Lash AE. CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res* **2007**; 35: D721-726.
[73] Afifi M. PubMed-indexed duplicate publications in the last decade, 1996-2006. *Ann Saudi Med* **2007**; 27: 302-304; author reply 304.
[74] Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* **2008**; 22: 338-342.
[75] Gass A, Doyle H. PLOS position on NIH public access policy. *Science* **2005**; 308: 356.
[76] Butler D. BioMed Central boosted by editorial board. *Nature* **2000**; 405: 384.
[77] Sabatti C, Rohlin L, Lange K, Liao JC. Vocabulon: a dictionary model approach for reconstruction and localization of transcription factor binding sites. *Bioinformatics* **2005**; 21: 922-931.
[78] Ahren DG, Ouzounis CA. Robustness of metabolic map reconstruction. *J Bioinform Comput Biol* **2004**; 2: 589-593.
[79] Karp PD, Paley S, Romero P. The Pathway Tools software. *Bioinformatics* **2002**; 1: S225-232.
[80] Nikitin F, Rance B, Itoh M, Kanehisa M, Lisacek F. Using protein motif combinations to update KEGG pathway maps and orthologue tables. *Genome Inform* **2004**; 15: 266-275.
[81] Paley SM, Karp PD. The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Res* **2006**; 34: 3771-3778.
[82] Ahn GT, Kim JH, Hwang EY, Lee MJ, Han IS. SCOPExplorer: a tool for browsing and analyzing structural classification of proteins (SCOP) data. *Mol Cells* **2004**; 17: 360-364.
[83] Chi PH, Shyu CR, Xu D. A fast SCOP fold classification system using content-based E-Predict algorithm. *BMC Bioinformatics* **2006**; 7: 362.
[84] Getz G, Starovolsky A, Domany E. F2CS. FSSP to CATH and SCOP prediction server. *Bioinformatics* **2004**; 20: 2150-2152.
[85] Gewehr JE, Hintermair V, Zimmer R. AutoSCOP. automated prediction of SCOP classifications using unique pattern-class mappings. *Bioinformatics* **2007**; 23: 1203-1210.
[86] Lo Conte L, Ailey B, Hubbard TJ, Brenner SE, Murzin AG, Chothia C. SCOP. a structural classification of proteins database. *Nucleic Acids Res* **2000**; 28: 257-259.
[87] Suhrer SJ, Wiederstein M, Sippl MJ. QSCOP--SCOP quantified by structural relationships. *Bioinformatics* **2007**; 23: 513-514.
[88] Tung CH, Yang JM. fastSCOP. a fast web server for recognizing protein structural domains and SCOP superfamilies. *Nucleic Acids Res* **2007**; 35: W438-443.
[89] Andreeva A, Howorth D, Brenner SE, Hubbard TJ, Chothia C, Murzin AG. SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res* **2004**; 32: D226-229.
[90] Andreeva A, Howorth D, Chandonia, *et al.* Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res* **2008**; 36: D419-425.
[91] Greene LH, Lewis TE, Addou S, *et al.* The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res* **2007**; 35: D291-297.
[92] Orengo CA, Pearl FM, Thornton JM. The CATH domain structure database. *Methods Biochem Anal* **2003**; 44: 249-271.
[93] Huynen MA, Snel B, Bork P, Gibson TJ. The phylogenetic distribution of frataxin indicates a role in iron-sulfur cluster protein assembly. *Hum Mol Genet* **2001**; 10: 2463-2468.
[94] Genschel U. Coenzyme A biosynthesis: reconstruction of the pathway in archaea and an evolutionary scenario based on comparative genomics. *Mol Biol Evol* **2004**; 21: 1242-1251.
[95] Striepen B, Pruijssers AJ, Huang J, *et al.* Gene transfer in the evolution of parasite nucleotide biosynthesis. *Proc Natl Acad Sci USA* **2004**; 101: 3154-3159.
[96] Suhre K. Inference of gene function based on gene fusion events: the rosetta-stone method. *Methods Mol Biol* **2007**; 396: 31-42.
[97] Suhre K, Claverie JM. FusionDB. a database for in-depth analysis of prokaryotic gene fusion events. *Nucleic Acids Res* **2004**; 32: D273-276.
[98] Panina EM, Vitreschak AG, Mironov AA, Gelfand MS. Regulation of biosynthesis and transport of aromatic amino acids in low-GC Gram-positive bacteria. *FEMS Microbiol Lett* **2003**; 222: 211-220.
[99] Berger MF, Bulyk ML. Protein binding microarrays (PBMs) for rapid, high-throughput characterization of the sequence specificities of DNA binding proteins. *Methods Mol Biol* **2006**; 338: 245-260.
[100] Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW, 3rd, Bulyk ML. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat Biotechnol* **2006**; 24: 1429-1435.

[101] Bulyk ML. Computational prediction of transcription-factor binding site locations. *Genome Biol* **2003**; 5: 201.

[102] Bulyk ML. Discovering DNA regulatory elements with bacteria. *Nat Biotechnol* **2005**; 23: 942-944.

[103] Bulyk ML. Analysis of sequence specificities of DNA-binding proteins with protein binding microarrays. *Methods Enzymol* **2006**; 410: 279-299.

[104] Bulyk ML. DNA microarray technologies for measuring protein-DNA interactions. *Curr Opin Biotechnol* **2006**; 17: 422-430.

[105] Bulyk ML. Protein binding microarrays for the characterization of DNA-protein interactions. *Adv Biochem Eng Biotechnol* **2007**; 104: 65-85.

[106] Bulyk ML, McGuire AM, Masuda N, Church GM. A motif co-occurrence approach for genome-wide prediction of transcription-factor-binding sites in Escherichia coli. *Genome Res* **2004**; 14: 201-208.

[107] McCord RP, Bulyk ML. Functional trends in structural classes of the DNA binding domains of regulatory transcription factors. *Pac Symp Biocomput* **2008**; 441-452.

[108] Mukherjee S, Berger MF, Jona G, *et al.* Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat Genet* **2004**; 36: 1331-1339.

[109] Ostrin EJ, Li Y, Hoffman K, Liu J, Wang K, Zhang L, Mardon G, Chen R. Genome-wide identification of direct targets of the Drosophila retinal determination protein Eyeless. *Genome Res* **2006**; 16: 466-476.

[110] Long F, Liu H, Hahn C, Sumazin P, Zhang MQ, Zilberstein A. Genome-wide prediction and analysis of function-specific transcription factor binding sites. *In Silico Biol* **2004**; 4: 395-410.

[111] Gertz J, Fay JC, Cohen BA. Phylogeny based discovery of regulatory elements. *BMC Bioinformatics* **2006**; 7: 266.

[112] Gertz J, Riles L, Turnbaugh P, Ho SW, Cohen BA. Discovery, validation, and genetic dissection of transcription factor binding sites by comparative and functional genomics. *Genome Res* **2005**; 15: 1145-1152.

[113] Tran TT, Dam P, Su Z, *et al.* Operon prediction in Pyrococcus furiosus. *Nucleic Acids Res* **2007**; 35: 11-20.

[114] Ashraf SI, Kunes S. A trace of silence: memory and microRNA at the synapse. *Curr Opin Neurobiol* **2006**; 16: 535-539.

[115] Chin JW. Programming and engineering biological networks. *Curr Opin Struct Biol* **2006**; 16: 551-556.

[116] Ciesla J. Metabolic enzymes that bind RNA. yet another level of cellular regulatory network**?** *Acta Biochim Pol* **2006**; 53: 11-32.

[117] Davidson EA, Ellington AD. Synthetic RNA circuits. *Nat Chem Biol* **2007**; 3: 23-28.

[118] Lee CT, Risom T, Strauss WM. Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. *DNA Cell Biol* **2007**; 26: 209-218.

[119] Narayanaswamy R, Ellington AD. Engineering RNA-based circuits. *Handb Exp Pharmacol* **2006**; 423-445.

[120] Redei GP, Koncz C, Phillips JD. Changing images of the gene. *Adv Genet* **2006**; 56: 53-100.

[121] Thomas CM. Transcription regulatory circuits in bacterial plasmids. *Biochem Soc Trans* **2006**; 34: 1072-1074.

[122] Mellin JR, Goswami S, Grogan S, Tjaden B, Genco CA. A novel fur- and iron-regulated small RNA, NrrF, is required for indirect fur-mediated regulation of the sdhA and sdhC genes in Neisseria meningitidis. *J Bacteriol* **2007**; 189: 3686-3694.

[123] Tjaden B. Prediction of small, noncoding RNAs in bacteria using heterogeneous data. *J Math Biol* **2008**; 56: 183-200.

[124] Tjaden B, Goodwin SS, Opdyke JA, Guillier M, Fu DX, Gottesman S, Storz G. Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res* **2006**; 34: 2791-2802.

[125] Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol* **2003**; 50: 1111-1124.

[126] To CC, Vohradsky J. Supervised inference of gene-regulatory networks. *BMC Bioinformatics* **2008**, 9: 2.

[127] Hsu SD, Chu CH, Tsou AP, *et al.* miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res* **2008**; 36: D165-169.

[128] Carlson MR, Zhang B, Fang Z, Mischel PS, Horvath S, Nelson SF. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics* **2006**; 7: 40.

[129] D'Haeseleer P, Liang S, Somogyi R. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* **2000**; 16: 707-726.

[130] Greller LD, Somogyi R. Reverse engineers map the molecular switching yards. *Trends Biotechnol* **2002**; 20: 445-447.

[131] Maki Y, Takahashi Y, Arikawa Y, *et al.* An integrated comprehensive workbench for inferring genetic networks: voyagene. *J Bioinform Comput Biol* **2004**; 2: 533-550.

[132] Maki Y, Tominaga D, Okamoto M, Watanabe S, Eguchi Y. Development of a system for the inference of large scale genetic networks. *Pac Symp Biocomput* **2001**; 446-458.

[133] Kao KC, Yang YL, Boscolo R, Sabatti C, Roychowdhury V, Liao JC. Transcriptome-based determination of multiple transcription regulator activities in Escherichia coli by using network component analysis. *Proc Natl Acad Sci USA* **2004**; 641-646.

[134] Liao JC, Boscolo R, Yang YL, Tran LM, Sabatti C, Roychowdhury VP. Network component analysis: reconstruction of regulatory signals in biological systems. *Proc Natl Acad Sci USA* **2003**; 100: 15522-15527.

[135] Sabatti C, Rohlin L, Oh MK, Liao JC. Co-expression pattern from DNA microarray experiments as a tool for operon prediction. *Nucleic Acids Res* **2002**; 30: 2886-2893.

[136] Zou M, Conzen SD. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics* **2005**; 21: 71-79.

[137] Almaas E. Biological impacts and context of network theory. *J Exp Biol* **2007**; 210: 1548-1558.

[138] Hardy S, Robillard PN. Petri net-based method for the analysis of the dynamics of signal propagation in signaling pathways. *Bioinformatics* **2008**; 24: 209-217.

[139] Jagalur M, Kulp D. An information theoretic method for reconstructing local regulatory network modules from polymorphic samples. *Comput Syst Bioinformatics Conf* **2007**; 6: 133-143.

[140] Margolin AA, Califano A. Theory and limitations of genetic network inference from microarray data. *Ann N Y Acad Sci* **2007**; 1115: 51-72.

[141] Faith JJ, Hayete B, Thaden JT, *et al.* Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profile**s**. *PLoS Biol* **2007**; 5: e8.

[142] Hayete B, Gardner TS, Collins JJ. Size matters: network inference tackles the genome scale. *Mol Syst Biol* **2007**; 3: 77.

[143] Margolin AA, Wang K, Lim WK, Kustagi M, Nemenman I, Califano A. Reverse engineering cellular networks. *Nat Protoc* **2006**; 1: 662-671.

[144] Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **2004**; 14: 283-291.

[145] Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L. Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J Mol Biol* **2006**; 360: 213-227.

[146] Luscombe NM, Babu MM. GenCompass: a universal system for analysing gene expression for any genome. *Trends Biotechnol* **2004**; 22: 552-555.

[147] Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **2004**, 431: 308-312.

[148] Baranzini SE, Mousavi P, Rio J, *et al.* Transcription-based prediction of response to IFNbeta using supervised computational methods. *PLoS Biol* **2005**; 3: e2.

[149] Fernald GH, Knott S, Pachner A, *et al.* Genome-wide network analysis reveals the global properties of IFN-beta immediate transcriptional effects in humans. *J Immunol* **2007**; 178: 5076-5085.

[150] Wen X, Fuhrman S, Michaels GS, Carr DB, Smith S, Barker JL, Somogyi R. Large-scale temporal gene expression mapping of central nervous system development. *Proc Natl Acad Sci USA* **1998**; 95: 334-339.

[151] Wang J, Kraemer E. GFPE. gene-finding program evaluation. *Bioinformatics* **2003**; 19: 1712-1713.

[152] Yugi K, Nakayama Y, Kojima S, Kitayama T, Tomita M. A microarray data-based semi-kinetic method for predicting quantitative dynamics of genetic networks. *BMC Bioinformatics* **2005**; 6: 299.

[153] Stark J, Brewer D, Barenco M, Tomescu D, Callard R, Hubank M. Reconstructing gene networks: what are the limits? *Biochem Soc Trans* **2003**; 31: 1519-1525.

[154] Nilsson R, Bajic VB, Suzuki H, *et al.* Transcriptional network dynamics in macrophage activation. *Genomics* **2006**; 88: 133-142.

[155] Barenco M, Tomescu D, Brewer D, Callard R, Stark J, Hubank M. Ranked prediction of p53 targets using hidden variable dynamic modeling. *Genome Biol* **2006**, 7: R25.

[156] Darvish A, Najarian K. Prediction of regulatory pathways using mRNA expression and protein interaction data: application to identification of galactose regulatory pathway. *Biosystems* **2006**; 83: 125-135.

[157] Ceol A, Chatr-aryamontri A, Santonico E, Sacco R, Castagnoli L, Cesareni G. DOMINO. a database of domain-peptide interactions. *Nucleic Acids Res* **2007**; 35: D557-560.

[158] Cesareni G, Ceol A, Gavrila C, Palazzi LM, Persico M, Schneider MV. Comparative interactomics. *FEBS Lett* **2005**; 579: 1828-1833.

[159] Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G. MINT. the Molecular INTeraction database. *Nucleic Acids Res* **2007**; 35: D572-574.

[160] Hermjakob H, Montecchi-Palazzi L, Bader G, *et al*. The HUPO PSI's molecular interaction format--a community standard for the representation of protein interaction data. *Nat Biotechnol* **2004**; 22: 177-183.

[161] Kerrien S, Orchard S, Montecchi-Palazzi L, *et al*. Broadening the horizon--level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol* **2007**; 5: 44.

[162] Orchard S, Salwinski L, Kerrien S, *et al*. The minimum information required for reporting a molecular interaction experiment (MIMIx). *Nat Biotechnol* **2007**; 25: 894-898.

[163] Persico M, Ceol A, Gavrila C, Hoffmann R, Florio A, Cesareni G. HomoMINT. an inferred human network based on orthology mapping of protein interactions discovered in model organisms. *BMC Bioinformatics* **2005**; 6 Suppl 4: S21.

[164] Gong S, Yoon G, Jang I, Bolser D, *et al*. PSIbase: a database of Protein Structural Interactome map (PSIMAP). *Bioinformatics* **2005**; 21: 2541-2543.

[165] Kim WK, Bolser DM, Park JH. Large-scale co-evolution analysis of protein structural interlogues using the global protein structural interactome map (PSIMAP). *Bioinformatics* **2004**; 20: 1138-1150.

[166] Park D, Lee S, Bolser D, Schroeder M, Lappe M, Oh D, Bhak J. Comparative interactomics analysis of protein family interaction networks using PSIMAP (protein structural interactome map). *Bioinformatics* **2005**; 21: 3234-3240.

[167] Ji ZL, Chen X, Zhen CJ, *et al*. KDBI. Kinetic Data of Biomolecular Interactions database. *Nucleic Acids Res* **2003**; 31: 255-257.

[168] Koike A, Takagi T. PRIME. automatically extracted PRotein Interactions and Molecular Information databasE. *In Silico Biol* **2005**; 5: 9-20.

[169] Mak HC, Daly M, Gruebel B, Ideker T. CellCircuits: a database of protein network models. *Nucleic Acids Res* **2007**; 35: D538-545.

[170] Mewes HW, Dietmann S, Frishman D, *et al*. MIPS. analysis and annotation of genome information in 2007. *Nucleic Acids Res* **2008**; 36: D196-201.

[171] Warner GJ, Adeleye YA, Ideker T. Interactome networks: the state of the science. *Genome Biol* **2006**; 7: 301.

[172] Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* **2002**; 30: 303-305.

[173] Bettencourt-Dias M, Giet R, Sinka R, *et al*. Genome-wide survey of protein kinases required for cell cycle progression. *Nature* **2004**, 432: 980-987.

[174] Linding R, Jensen LJ, Pasculescu A, Olhovsky M, Colwill K, Bork P, Yaffe MB, Pawson T. NetworKIN. a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res* **2008**; 36: D695-699.

[175] Lim YP, Diong LS, Qi R, Druker BJ, Epstein RJ. Phosphoproteomic fingerprinting of epidermal growth factor signaling and anticancer drug action in human tumor cells. *Mol Cancer Ther* **2003**, 2: 1369-1377.

[176] Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, Jensen ON. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* **2005**; 4: 310-327.

[177] Nuhse TS, Stensballe A, Jensen ON, Peck SC. **P**hosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. *Plant Cell* **2004**; 16: 2394-2405.

[178] Peck SC. Phosphoproteomics in Arabidopsis: moving from empirical to predictive science. *J Exp Bot* **2006**; 57; 1523-1527.

[179] Wu J, Smith LT, Plass C, Huang TH. ChIP-chip comes of age for genome-wide functional analysis. *Cancer Res* **200**6; 66: 6899-6902.

[180] Jung GY, Stephanopoulos G. A functional protein chip for pathway optimization and *in vitro* metabolic engineering. *Science* **2004**; 304: 428-431.

[181] Kim M, Lee SY, Choi H, Shin YB, Jung SO, Kim MG, Chung BH. On-chip Escherichia coli culture, purification, and detection of expressed proteins. *Eur Biophys J* **2006**; 35: 655-662.

[182] Marino S, Voit EO. An automated procedure for the extraction of metabolic network information from time series data. *J Bioinform Comput Biol* **2006**, 4: 665-691.

[183] Veflingstad SR, Almeida J, Voit EO. Priming nonlinear searches for pathway identification. *Theor Biol Med Model* **2004**; 1: 8.

[184] Voit EO, Almeida J, Marino S, Lall R, Goel G, Neves AR, Santos H. Regulation of glycolysis in Lactococcus lactis: an unfinished systems biological case study. *Syst Biol (Stevenage)* 2006; 153: 286-298.

[185] Alvarez-Vasquez F, Sims KJ, Cowart LA, Okamoto Y, Voit EO, Hannun YA. Simulation and validation of modelled sphingolipid metabolism in Saccharomyces cerevisiae. *Nature* **2005**; 433: 425-430.

[186] Vance W, Arkin A, Ross J. Determination of causal connectivities of species in reaction networks. *Proc Natl Acad Sci USA* **2002**; 99: 5816-5821.

[187] Famili I, Forster J, Nielsen J, Palsson BO. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc Natl Acad Sci USA* **2003**; 100: 13134-13139.

[188] Becker SA, Palsson BO. Three factors underlying incorrect in silico predictions of essential metabolic genes. *BMC Syst Biol* **2008**, 2: 14.

[189] Chen KC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, Tyson JJ. Integrative analysis of cell cycle control in budding yeast. *Mol Biol Cell* **2004**; 15: 3841-3862.

[190] Cross FR, Archambault V, Miller M, Klovstad M. Testing a mathematical model of the yeast cell cycle. *Mol Biol Cell* **2002**; 13: 52-70.

[191] Cross FR, Siggia ED. Mode locking the cell cycle. *Phys Rev E Stat Nonlin Soft Matter Phys* **2005**; 72: 021910.

[192] Boros LG, Brandes JL, Lee WN, *et al*. Thiamine supplementation to cancer patients: a double edged sword. *Anticancer Res* **1998**; 18: 595-602.

[193] Boros LG, Brandes JL, Yusuf FI, Cascante M, Williams RD, Schirmer WJ. Inhibition of the oxidative and nonoxidative pentose phosphate pathways by somatostatin: a possible mechanism of antitumor action. *Med Hypotheses* **1998**; 50: 501-506.

[194] Boros LG, Lee PW, Brandes JL, *et al*. Nonoxidative pentose phosphate pathways and their direct role in ribose synthesis in tumors: is cancer a disease of cellular glucose metabolism? *Med Hypotheses* **1998**; 50: 55-59.

[195] Cascante M, Boros LG, Comin-Anduix B, de Atauri P, Centelles JJ, Lee PW. Metabolic control analysis in drug discovery and disease. *Nat Biotechnol* **2002**; 20: 243-249.

[196] Girgis S, Pai SM, Girgis IG, Batra VK. Pharmacodynamic parameter estimation: population size versus number of samples. *AAPS J* **2005**; 7: 46.

[197] Beltz WF. Estimation and use of kinetic parameter distributions in metabolism and nutrition. *Adv Food Nutr Res* **1996**; 40: 265-279.

[198] Tyukin I, van Leeuwen C, Prokhorov D. Parameter estimation of sigmoid superpositions: dynamical system approach. *Neural Comput* **2003**; 15: 2419-2455.

[199] Hautaniemi S, Kharait S, Iwabu A, Wells A, Lauffenburger DA. Modeling of signal-response cascades using decision tree analysis. *Bioinformatics* **2005**; 21: 2027-2035.

[200] Kharait S, Hautaniemi S, Wu S, Iwabu A, Lauffenburger DA, Wells A. Decision tree modeling predicts effects of inhibiting contractility signaling on cell motility. *BMC Syst Biol* **2007**; 1: 9.

[201] Chance B, Greenstein DS, Roughton FJ. The mechanism of catalase action. I. Steady-state analysis. *Arch Biochem* **1952**; 37: 301-321.

[202] Sorribas A, Hernandez-Bermejo B, Vilaprinyo E, Alves R. Cooperativity and saturation in biochemical networks: A saturable formalism using Taylor series approximations. *Biotechnol Bioeng* **2006.**

[203] Michaelis L, Menten M. Die Kinetik der Invertinwirkung. *Biochem Z* **1913**; 49: 333-369.
[204] Alves R, Vilaprinyo E, Hernandez-Bermejo B, Sorribas A. Mathematical formalisms based on approximated kinetic representations for modeling genetic and metabolic pathways. *Biotechnology and Genetic Engineering Reviews* **2008**; **25**: submitted.
[205] Sorribas A, Hernandez-Bermejo B, Vilaprinyo E, Alves R. Cooperativity and saturation in biochemical networks: a saturable formalism using Taylor series approximations. *Biotechnol Bioeng* **2007**; 97: 1259-1277.
[206] Ideker T. A systems approach to discovering signaling and regulatory pathways--or, how to digest large interaction networks into relevant pieces. *Adv Exp Med Biol* 2004; 547: 21-30.
[207] Ideker T, Lauffenburger D. Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol* **2003**; 21: 255-262.
[208] Ideker T, Valencia A. Bioinformatics in the human interactome project. *Bioinformatics* **2006**; 22: 2973-2974.
[209] Su Z, Dam P, Chen X, Olman V, Jiang T, Palenik B, Xu Y. Computational inference of regulatory pathways in microbes: an application to phosphorus assimilation pathways in Synechococcus sp. WH8102. *Genome Inform* **2003,** 14: 3-13.
[210] Su Z, Mao F, Dam P, Wu H, Olman V, Paulsen IT, Palenik B, Xu Y. Computational inference and experimental validation of the nitrogen assimilation regulatory network in cyanobacterium Synechococcus sp. WH. *Nucleic Acids Res* **2006**; **34**: 1050-1065.
[211] Gutierrez-Rios RM, Rosenblueth DA, Loza JA, *et al.* Regulatory network of Escherichia coli: consistency between literature knowledge and microarray profiles. *Genome Res* **2003**; 13: 2435-2443.
[212] Wang L, Yang Z, Li Y, *et al.* Reconstruction and in silico analysis of the MAPK signaling pathways in the human blood fluke, Schistosoma japonicum. *FEBS Lett* **2006**; 580: 3677-3686.
[213] Ingham RJ, Colwill K, Howard C, *et al*. WW domains provide a platform for the assembly of multiprotein networks. *Mol Cell Biol* **2005**; 25: 7092-7106.
[214] Linding R, Jensen LJ, Ostheimer GJ, *et al*. Systematic discovery of *in vivo* phosphorylation networks. *Cell* **2007**; 129: 1415-1426.
[215] Forster AC, Church GM. Synthetic biology projects *in vitro*. *Genome Res* **2007, 17**: 1-6.
[216] Greber D, Fussenegger M. Mammalian synthetic biology: engineering of sophisticated gene networks. *J Biotechnol* **2007**; 130: 329-345.
[217] Kaznessis YN. Models for synthetic biology. *BMC Syst Biol* **2007**; 1: 47.
[218] Keasling JD. Synthetic biology for synthetic chemistry. *ACS Chem Biol*, **2008**; 3: 64-76.
[219] Pleiss J. The promise of synthetic biology. *Appl Microbiol Biotechnol* **2006**; 73: 735-739.
[220] Saito H, Inoue T. RNA and RNP as new molecular parts in synthetic biology. *J Biotechnol* **2007**; 132: 1-7.
[221] Sole RV, Munteanu A, Rodriguez-Caso C, Macia J. Synthetic protocell biology: from reproduction to computation. *Philos Trans R Soc Lond B Biol Sci* **2007**; 362: 1727-1739.
[222] Aslett M, Wood V. Gene Ontology annotation status of the fission yeast genome: preliminary coverage approaches 100%. *Yeast* **2006**; 23: 913-919.
[223] Azuaje F, Al-Shahrour F, Dopazo J. Ontology-driven approaches to analyzing data in functional genomics. *Methods Mol Biol* **2006**; 316: 67-86.
[224] Dybkaer R. An ontology on property for physical, chemical, and biological systems. *APMIS Suppl* **2004**: 1-210.
[225] Jenkins H, Hardy N, Beckmann M, *et al*. A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol* **2004**, 22: 1601-1606.
[226] Lefranc MP. IMGT-ONTOLOGY and IMGT databases, tools and Web resources for immunogenetics and immunoinformatics. *Mol Immunol* **2004**; 40: 647-660.
[227] Lomax J. Get ready to GO! A biologist's guide to the Gene Ontology. *Brief Bioinform* **2005**, 6: 298-304.
[228] Shapiro JS, Bakken S, Hyun S, Melton GB, Schlegel C, Johnson SB. Document ontology: supporting narrative documents in electronic health records. *AMIA Annu Symp Proc* **2005**; 684-688.
[229] Snowden DJ. Multi-ontology sense making: a new simplicity in decision making. *Inform Prim Care* **2005**; 13: 45-54.
[230] Whetzel PL, Brinkman RR, Causton HC, *et al*. Development of FuGO. an ontology for functional genomics investigations. *Omics* **2006**; 10: 199-204.
[231] Whetzel PL, Parkinson H, Stoeckert CJ, Jr.: Using ontologies to annotate microarray experiments. *Methods Enzymol* **2006**; 411: 325-339.
[232] Yu AC. Methods in biomedical ontology. *J Biomed Inform* **2006**; 3**9**: 252-266.
[233] Yue L, Reisdorf WC. Pathway and ontology analysis: emerging approaches connecting transcriptome data and clinical endpoints. *Curr Mol Med* **2005**; 5: 11-21.
[234] Myers CL, Barrett DR, Hibbs MA, Huttenhower C, Troyanskaya OG. Finding function: evaluation methods for functional genomic data. *BMC Genomics* **2006**; **7**: 187.
[235] Jones AR, Pizarro A, Spellman P, Miller M. FuGE. Functional Genomics Experiment Object Model. *Omics* **2006**; 10: 179-184.
[236] Le Novere N, Finney A, Hucka M, *et al*. Minimum information requested in the annotation of biochemical models (MIRIAM**)**. *Nat Biotechnol* **2005**; 23: 1509-1515.
[237] Spellman PT, Miller M, Stewart J, *et al*. Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol* **2002**; 3: RESEARCH0046.
[238] Chervitz SA, Hester ET, Ball CA, *et al*. Using the Saccharomyces Genome Database (SGD) for analysis of protein similarities and structure. *Nucleic Acids Res* **1999**; 27: 74-78.
[239] Christie KR, Weng S, Balakrishnan R, *et al*. Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms. *Nucleic Acids Res* **2004**, 32: D311-314.
[240] Dwight SS, Harris MA, Dolinski K, *et al.* Saccharomyces Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucleic Acids Res* **2002**; 30: 69-72.
[241] Ostell JM, Wheelan SJ, Kans JA. The NCBI data model. *Methods Biochem Anal* **2001**; 43: 19-43.
[242] Shah SP, Huang Y, Xu T, Yuen MM, Ling J, Ouellette BF. Atlas - a data warehouse for integrative bioinformatics. *BMC Bioinformatics* 2005; 6: 34.
[243] Weng S, Dong Q, Balakrishnan R, *et al.* Saccharomyces Genome Database (SGD) provides biochemical and structural information for budding yeast proteins. *Nucleic Acids Res* **2003**; 31: 216-218.
[244] Kolinski A, Bujnicki JM. Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. *Proteins* **2005**; 61 Suppl **7**: 84-90.
[245] Lundstrom J, Rychlewski L, Bujnicki J, Elofsson A. Pcons: a neural-network-based consensus predictor that improves fold recognition. *Protein Sci* **2001**; 10: 2354-2362.
[246] Sen TZ, Kloster M, Jernigan RL, Kolinski A, Bujnicki JM, Kloczkowski A. Predicting the complex structure and functional motions of the outer membrane transporter and signal transducer FecA. *Biophys J* **2008.**
[247] Bujnicki JM, Elofsson A, Fischer D, Rychlewski L. Structure prediction meta server. *Bioinformatics* **2001**; 17: 750-751.
[248] Eyrich VA, Rost B. META-PP. single interface to crucial prediction servers. *Nucleic Acids Res* **2003**; 31: 3308-3310.
[249] Grana O, Eyrich VA, Pazos F, Rost B, Valencia A. EVAcon: a protein contact prediction evaluation service. *Nucleic Acids Res* **2005**; 33: W347-351.
[250] Koh IY, Eyrich VA, Marti-Renom MA, A *et al.* EVA. Evaluation of protein structure prediction servers. *Nucleic Acids Res* **2003**; 31: 3311-3315.
[251] Fischer D. Servers for protein structure prediction. *Curr Opin Struct Biol* **2006**; **16**: 178-182.
[252] Friedberg I, Harder T, Godzik A. JAFA. a protein function annotation metaserver. *NucleicAcids Res* 2006, **34**: W379-381
[253] Geueke M, Stausberg J. A meta-data-based learning resource server for medicine. *Comput Methods Programs Biomed* **2003**; 72: 197-208.
[254] Ginalski K, Pas J, Wyrwicz LS, von Grotthuss M, Bujnicki JM, Rychlewski L. ORFeus: Detection of distant homology using sequence profiles and predicted secondary structure. *Nucleic Acids Res* **2003**; 31: 3804-3807.
[255] von Grotthuss M, Pas J, Wyrwicz L, Ginalski K, Rychlewski L. Application of 3D-Jury, GRDB, and Verify3D in fold recognition. *Proteins* **2003**; 53 Suppl 6: 418-423.
[256] Kosinski J, Cymerman IA, Feder M, Kurowski MA, Sasin JM, Bujnicki JM. A "FRankenstein's monster" approach to comparative

modeling: merging the finest fragments of Fold-Recognition models and iterative model refinement aided by 3D structure evaluation. *Proteins* **2003**; 53 Suppl 6: 369-379.

[257] Kosinski J, Gajda MJ, Cymerman IA, *et al.* FRankenstein becomes a cyborg: the automatic recombination and realignment of fold recognition models in CASP6. *Proteins* **2005**; 61 Suppl 7: 106-113.

[258] Kurowski MA, Bujnicki JM. GeneSilico protein structure prediction meta-server. *Nucleic Acids Res* **2003**; 31: 3305-3307.

[259] Sasin JM, Kurowski MA, Bujnicki JM. STRUCLA. a WWW meta-server for protein structure comparison and evolutionary classification. *Bioinformatics* **2003,** 19 Suppl 1: i252-254.

[260] Lemkin PF. The 2DWG meta-database of two-dimensional electrophoretic gel images on the Internet. *Electrophoresis* **1997**; 18: 2759-2773.

[261] Saini HK, Fischer D. Meta-DP. domain prediction meta-server. *Bioinformatics* **2005**; 21: 2917-2920.

[262] Saini HK, Fischer D. FRalanyzer: a tool for functional analysis of fold-recognition sequence-structure alignments. *Nucleic Acids Res* **2007**; 35: W499-502.

[263] von Grotthuss M, Koczyk G, Pas J, Wyrwicz LS, Rychlewski L. Ligand.Info small-molecule Meta-Database. *Comb Chem High Throughput Screen* **2004, 7**: 757-761.

[264] Wallner B, Elofsson A. Pcons5: combining consensus, structural evaluation and fold recognition scores. *Bioinformatics* **2005**; 21: 4248-4254.

[265] Wallner B, Elofsson A. Prediction of global and local model quality in CASP7 using Pcons and ProQ. *Proteins* **2007**; 69 Suppl 8: 184-193.

[266] Wallner B, Larsson P, Elofsson A. Pcons.net: protein structure prediction meta server. *Nucleic Acids Res* **2007**; 35: W369-374.

[267] Carrere S, Gouzy J. REMORA. a pilot in the ocean of BioMoby web-services. *Bioinformatics* **2006**, 22: 900-901.

[268] Kawas E, Senger M, Wilkinson MD. BioMoby extensions to the Taverna workflow management and enactment software. *BMC Bioinformatics* **2006**; **7**: 523.

[269] Wilkinson M. Gbrowse Moby: a Web-based browser for BioMoby Services. *Source Code Biol Med* **2006**; 1: 4.

[270] Wilkinson M, Schoof H, Ernst R, Haase D. BioMOBY successfully integrates distributed heterogeneous bioinformatics Web Services. The PlaNet exemplar case. *Plant Physiol* **2005**; 138: 5-17.

[271] Wilkinson MD, Links M. BioMOBY. an open source biological web services proposal. *Brief Bioinform* 2002, 3: 331-341.

[272] Stevens RD, Robinson AJ, Goble CA. myGrid: personalised bioinformatics on the information grid. *Bioinformatics* **2003**; 19: i302-304.

[273] Covitz PA, Hartel F, Schaefer C, *et al.* caCORE. a common infrastructure for cancer informatics. *Bioinformatics* **2003**; 19: 2404-2412.

[274] Feigenbaum L, Herman I, Hongsermeier T, Neumann E, Stephens S. The Semantic Web in action. *Sci Am* **2007**; 297: 64-71.

[275] Dhanapalan L, Chen JY. A case study of integrating protein interaction data using semantic web technology. *Int J Bioinform Res Appl* **2007**; 3: 286-302.

[276] Neumann E, Prusak L. Knowledge networks in the age of the Semantic Web. *Brief Bioinform* **2007**; 8: 141-149.

[277] Post LJ, Roos M, Marshall MS, van Driel R, Breit TM. A semantic web approach applied to integrative bioinformatics experimentation: a biological use case with genomics data. *Bioinformatics* **2007**; 23: 3080-3087.

[278] Quan D. Improving life sciences information retrieval using semantic web technology. *Brief Bioinform* **2007**, 8: 172-182.

[279] Ruttenberg A, Clark T, Bug W, *et al.* Advancing translational research with the Semantic Web. *BMC Bioinformatics* 2007, 8: S2.

[280] Smith A, Cheung K, Krauthammer M, Schultz M, Gerstein M. Leveraging the structure of the Semantic Web to enhance information retrieval for proteomics. *Bioinformatics* **2007**; 23: 3073-3079.

[281] Smith AK, Cheung KH, Yip KY, Schultz M, Gerstein MK. LinkHub: a Semantic Web system that facilitates cross-database queries and information retrieval in proteomics. *BMC Bioinformatics* **2007**; 3: S5.

SUPPLEMENTARY APPENDIX: AVAILABLE DATASETS, ANALYTICAL METHODS AND TOOLS

Here we provide a more in depth analysis of the different datasets that are available for metabolic reconstruction and of the tools that can be used to mine those data sets for information.

S.1 Literature Data

Available Datasets

Because thousands of scientific papers are published monthly within many areas of molecular biology, it is not feasible for anyone to read them all. A digest of the information contained in the literature is available on-line, in literature databases that contain the abstract, title and keyword information such as ***MEDLINE*** [1], ***PUBMED*** [2], ***Web of Science*** [3, 4] or others [5]. The full text of many papers and some books is also available in databases available in journal home pages. The information in these databases is organized according to strict criteria that facilitate finding papers reporting on many different subjects. The various databases are associated to diverse search engines that have different sensitivities and goals [6-11]. Automated analysis of a natural text within a given literature database to find specific pieces of information (gene names, gene function, parameter values, interaction information between individual molecular components, etc.) is a rapidly evolving field [12]. Truly flexible, multipurpose, and automated network mining from the full text of live documents brings enormous added value from the literature datasets to molecular systems biology research [13-16]. Values for kinetic and thermodynamic parameters for many different reactions and processes are also buried in the literature. Mining for this data has been mostly manual, and parameter information stored in databases is not in general sufficient to evaluate the usability of such data (e.g. ***BRENDA*** and enzyme parameters [17-19]).

Analytical Methods and Tools for Text Mining

Natural language processing and mining is developing rapidly [3, 6, 20-39]. Currently, several alternative models are being used to develop tools that mine the literature. One of the models is ontology-based and relies on creating classifications that reduce ambiguity in the text to be mined (e.g. [3, 22, 25]). Another model is based upon the existence of thesaurus and dictionaries of synonyms [26, 40]. A third model, based on machine learning and using a combination of ontologies and thesaurus, allows the search algorithm to be trained while mining text information (e. g. [20, 21]). This should improve the performance of the methods in a near future. As a general goal, it is increasingly important to make the methods for automated text mining available to the biological researchers in applications that facilitate the derivation of biologically relevant information. Some text mining tools that fulfill those requirements are already available (e. g. [20, 36, 40-48]). For example, they can be used to extract network information [41, 42, 44, 47] or protein localization [48] from raw text and literature analysis. It would also be invaluable to develop specific software applications that allow researchers to accurately identify parameter value reports. Ideally, an automated mining of parameter information would allow for the identification of organism, reaction, units and conditions of parameter determination.

It is likely that the major hindrance to a precise and complete mining of the literature information by automated text mining tools will be the decades of accumulated literature using non-standardized nomenclature and classifications (see main text). Possible remediation for these problems are a) creating better dictionaries of synonyms, and b) using artificial intelligence and machine learning methods to increase the size of the dictionaries and allow the algorithms to make educated guesses regarding correspondence of terms. Part of this upgrading of the dictionaries could be done according to a combination of machine learning and a "wiki-like" model. That is, as users executed their jobs they could provide information and feedback to the methods regarding new synonyms and correct/incorrect guesses made by the software.

S.2 Sequence Data

Available Datasets

With the increasing number of genome sequencing projects that are completed or underway, increasing amounts of sequence data are available. Sequence datasets are too numerous to mention. However, a few central repositories contain most of the sequence information, often duplicated among the repositories. The ***NCBI*** web pages [36-42] and the ***KEGG*** web pages [43-47] contain annotation information for most organisms with fully sequenced genomes. ***Expasy*** [48, 49] also has this information with a focus on the proteins. The ***TIGR*** institute [50, 51] web pages and the ***DOE Joint Genome Institute*** [52] web pages also contain much of this information. Then, individual genome project web pages contain a wealth of functional details regarding their genome of interest and its genes. This is the case for example of the ***SGD*** resource [53, 54] for the yeast ***Saccharomyces cerevisiae***, of WormBase for *C. elegans* [49] or ***FlyBase*** for *Drosophila* [50, 51]. There are also databases available for RNA genes and gene targets (e. g. [52-55]) as well as for regulatory sequences in the genome of different species (e. g. [56]).

Analytical Methods and Tools for Functional Annotation of Sequences

We will not discuss the computational tools and methods used for genome sequencing and annotation. Those interested can find detailed explanations of the historical evolution of genome sequencing methods in the paper by Ankeny [57] and the references therein. For more recent reviews on the subject see [58-61].

Once the genome sequence is completed and assembled, one faces the task of finding the open reading frames (ORFs). Searching for homology to known genes is one of the methods for genome annotation. The *de facto* standard tool for aligning homologous sequences is the ***BLAST*** alignment program [62]. When no sufficient homology can be found, other methods and tools exist to predict ORFs. An overview of such methods can be found in [63]. By and large, the gene finding methods split into signal finding methods or content finding methods. Signal finding methods look for elements that signal the beginning and the end of genes, such as the ATG codon, or sequences that are recog-

nized as promoter sequences. However, many other signals that can be used for this purpose exist in genomes. These signals include for example transcription factor binding sites or ribosome binding sites in prokaryotes. Content finding methods are based upon finding differences in general statistical properties between coding and non-coding sequences of the genome. For example, in a genome the frequency of dinucleotides is different in coding and non-coding sequences, as is the GC content of the sequence. These differences can be used to separate the two types of sequences. Different gene finding tools use alternative combinations of methods to find signals and content, thus improving their probability of finding true ORFs [64-70]. The available tools for gene finding include ***GENEMARK*** [64], ***SNAP*** [68], ***Ensembl*** [71], ***SPG2*** or ***TWINSCAN*** (e.g. [67]), among many others.

Metagenomic sequencing efforts, where the DNA sequence of an environment is sequenced and must be assembled, are now emerging [72-76]. Finding genes in these sequences is not easy because often one does not know to which genome a sequence belongs to. Ideally, the sequences should be assembled, whenever possible, into its individual genomes which can be analyzed to find genes. However, if such assembly is not possible, specialized programs must be used to predict genes in large fractions of orphan DNA [77, 78]. ***Metagene*** [69] is an example of a tool developed specifically for prokaryotic gene finding in metagenomic samples.

Several centralized servers such as ***Expasy*** [79], ***KEGG*** [80-83] or ***NCBI*** [84] provide tools that allow researchers to infer functional information from sequence data. The information made available for any given sequence ranges from the actual molecular function of an individual gene/protein to the prediction of post translational modifications in a protein [77] or the prediction of its location within the cell [85-88]. Also, several on-line services provide maps of metabolic pathways, gene circuits, and signal transduction pathways [89-100]. By superimposing the different genes identified within a genome onto those maps, one gains knowledge about qualitative aspects of cellular behavior. For example, if there are no genes coding for a given amino acid pathway within a genome one can expect that the organism will require that amino acid as a part of its diet. Some tools and servers that allow for analysis of phylogenetic conservation and congruence studies between genes in sets of fully sequenced genomes are also available [101-108]. There are also tools available for the analysis of RNA genes and for the analysis of regulation by RNA [109-115].

S.3 Structural Data: Mining for Function

Available Datasets

As is the case for sequence data, the amount of structural data for the molecular components of cells is also increasing steadily. The whole proteome structural genomics projects [116, 117] that are currently under way contribute to this increase [118]. There is a central repository of protein and nucleic acid structures at the protein databank (***PDB***) [119]. The structures deposited at the ***PDB*** include complexes formed between proteins and even provide information regarding the stoichiometry of those complexes. Many of the structures contained in the ***PDB*** can also be found elsewhere, for example at the ***NCBI*** [84].

Analytical Methods and Tools

Computational methods and tools are necessary from the ground up in the analysis of structural data. The most common ways to obtain protein structure are:

a) Submitting purified protein crystals to high energy radiation bombardment (e. g. X-Ray or neutrons) and capturing snapshots of the radiation diffracted by the protein crystals. These snapshots are then reconstituted into a 3D structure using appropriate software [120-125].

b) Using NMR technology to take spectral snapshots of the protein in solution [126-130], that are then reconstituted into a 3D structure using appropriate software.

Technical details regarding structure determination are well beyond the scope of this review. For practical reasons, our main focus will be on structure visualization, analysis and prediction. Once the proteins structures are available they are reported and stored using a specific data format [131-133] that includes, among other things, the 3D coordinates of the different atoms in the protein, the native organism, and the oligomerization state of the biologically relevant form of the protein. An analysis of these structures requires tools for their visualisation. One of the simplest tools available is ***RasMol*** [134]. This free tool allows the researcher to visualize and do minimal manipulation of the structure files. Another free tool is ***DEEPVIEW*** [135]. This tool is more sophisticated and, for example, it allows for point mutations to be introduced into the structure and energy minimization to be performed. It also allows researcher to do their own structural modelling of their protein of interest if the structure has not been determined yet. There are alternative methods for structural modelling. The most successful can be roughly grouped as follows:

a) **Fold prediction.** In some cases, when not enough information is available to obtain a 3D structure prediction, there are methods that allow the prediction of the secondary structure of a protein [136, 137]. Currently, many of the methods used for fold prediction are used as intermediate steps for obtaining a 3D structure prediction of proteins for which there are no known homologues in the PDB.

b) **Homology modelling.** The researcher has a sequence of interest and finds homologues of that sequence that have had their structure determined. By doing a sequence alignment, the researcher can then superimpose the sequence onto the known structure, followed by energy optimization. Many servers do homology modelling. ***SWISSMODEL*** [135] and ***3DJIGSAW*** [138] are a few of these. To find a list of other such servers one can consult the special issues of the journal Proteins that report the ***CASP*** (critical assessment of structural prediction) results [139-145].

c) **Threading.** In the absence of homologues with known structure, one can thread the sequence over a known structure of a protein that one thinks is likely to have structural similarity, followed by energy minimization.

d) ***Ab initio* modelling.** By using short stretches of homology over different protein structures one can also create models for the structure of a protein that has no homologues with a known structure. ***Robetta*** is an example of a server that allows researchers to do *ab initio* modelling from sequence [146]. To find a list of other such servers see the special issues of the journal Proteins that report the ***CASP*** results [139-145]. An alternative *ab initio* method is starting with the linear alpha helical structure of a sequence and by doing molecular dynamics simulations study how the protein folds. However, there is not enough computational power available to use this approach as a standard procedure.

The use of *in silico* protein docking can also reveal details about how different biological molecules physically interact. Interested readers can use different programs and servers to perform such docking studies [147-164].

S.4 Gene Expression Data

Available Data

Large datasets measuring changes in gene expression for entire genomes, under different conditions are accessible for many different organisms. These data can be used to infer which genes and proteins are important for a specific cellular response, thus giving researchers information about the role of genes with unknown function. Additionally one can infer information regarding the dynamics of the function, which is something harder, or even impossible, to do from sequence or structural data. Gene expression data is deposited in different databases. The Gene Expression Omnibus ***(GEO),*** at the ***NCBI*** offers the raw data for many of the micro array experiments published in the literature [165]. Other general databases exist, both for microarray data and for other types of gene expression experiments (e. g. [166-171]). There are also databases for gene expression for specific human diseases (e.g. [172, 173]) in specific organisms, such as mouse [174], *Arabidopsis* [175] and others.

Analytical Methods and Tools

With the advent of fully sequenced genomes, there came the invention of DNA chips and microarrays [176, 177]. In microarray experiments, probes that hybridize mRNA for a large fraction of the individual genes in a genome are imprinted onto a slab of material. Cells are grown and subject to some stressful condition, which may include mutation. Then cells are collected and their mRNAs are purified, amplified and hybridized with the probes in the microarray. By comparison with a control condition, these experiments reveal how much the expression of individual genes changed in response to the stress. For a more detailed description of microarray experiments see for example [178].

Other methods also allow measuring changes in gene expression at the whole genome level. Such methods include for example SAGE [179, 180] or CAGE [181] experiments, where mRNAs are directly amplified and identified in solution without the need of an imprinted microarray. A variation of the classical DNA microarray is the tiling microarray. The probes in these arrays are designed to cover entire regions of the genome, rather than individual genes [182]. These microarrays have shown that there are large islands of expression, for example in the human genome, that are either outside the genes or constitute only fragments of those genes [182, 183].

As is the case with the determination of protein structure, computational methods are needed to process the data in microarray experiments that measure changes in gene expression. The determination of change fold for gene expression in microarray experiments is usually done using differential measurements of fluorescence or radioactivity. Image capture, processing and noise filtering are areas where appropriate algorithms must be used in order to ensure acceptable data acquisition and normalization [178, 184, 185].

Once the data are collected and organized in its final form, then a different set of computational methods is required in order to determine whether a change in gene expression is actually significant or whether it can be attributed to noise [186-189]. There are no general methods that are universally considered as more appropriate to analyze the significance of a given change fold. This is an area of intensive research. Microarray data have problems of reproducibility and robustness, which underlines the importance of experimental design, replicate experiments and appropriate statistical methods as a means to identify significant changes in gene expression.

There are several tools to analyze microarray gene expression data and cluster genes according to their changes in expression. For example, the tools provided in ***GEO*** [165, 190] can be used to comparing different gene expression profiles. ***TM4*** [191] and ***BASE*** [192] are examples of software platforms that allow for the analysis and comparison of gene expression profiles. ***Bioconductor*** [193] is a popular open source software platform that uses a set of functions and procedures written in R to analyze microarray data.

S.5 Proteomics Data

Available Data

What information from proteomics experiments should be stored in databases and how it should be organized is still not consensual [194-200]. In consequence there is still no central repository of proteomics data. ***PRIDE*** [199, 201] and the ***Peptide Atlas*** [202] projects are probably the resources that contain more data from proteomics experiments. The global proteomics machine [203] provides software and internet services that mine data from these and other proteomics data sources. On a smaller scale, the ***NetworKIN*** database stores proteomics information about the phosphorylation state of human phosphoproteins [204].

There is, to our knowledge, no good and general HTP method to characterize protein activity on a large scale (however, see the main text for a promising new method.) The development of such methods would allow for the large scale determination of enzyme parameters, binding, and

thermodynamics constants and would likely create a revolution in mathematical modeling. Nevertheless, and independently of the limitations, the application of proteomics methods to the study of the dynamics of protein post translational modification can be a step forward in obtaining this type of information.

Analytical Methods and Tools

The progresses in 2D gel electrophoresis and in Mass Spectrometry (MS) and NMR technology have permitted the analysis of whole cell samples in order to identify the protein complement of the cell [205-225]. In general, proteins are separated via electrophoresis. The different proteins are then identified using MS techniques or NMR techniques. Unfortunately, datasets from these studies are not as abundant nor publicly available as those from microarray experiments. Furthermore, a central repository for all the proteomics experimental data is still lacking. Nevertheless, the available proteomics data already provide information regarding community proteomics [226], subcellular localization of proteins [227-229], dynamics of protein turnover [207, 208, 230, 231], post-translational modifications of proteins [217, 232-240], markers of specific responses, and many other aspects of protein function.

Most of the software effort in proteomics is still dedicated to creating tools for identifying proteins from the MS or NMR spectra. These tools work, for example, by using Fourier analysis for spectral identification. The spectra are compared to know protein spectra, thus allowing for the identification of proteins in sample. The data sources mentioned in the previous section provide tools that allow researchers to identify the proteins that are present or absent in each sample.

S.6 Metabolomics Data

Available Data

There is, to the knowledge of the authors, no general central public repository of metabolomics data. A few databases already exist, but their content is fragmentary, with respect to the amount of reported metabolomics data [241-244]. Furthermore, the problem of developing a standard in data reporting is a difficult one that has only now began to be addressed [245-249]. Links to some of the current repositories of metabolomics data available on the web can be found at http://www.bmrb.wisc.edu/metabolomics/external_metab_links.html. As time progresses and standardization settles into the field, it is likely that central repositories of metabolomics data will be developed. Such data will be invaluable for example to derive parameter values and perform mathematical model validation [250].

Analytical Methods and Tools

Cells use proteins to take up nutrients and make other small molecular species. There may be more than 200 000 metabolites in the plant kingdom alone [200]. Unlike proteomics or genomics, in metabolomics experiments hundreds of types of molecules with different chemical properties must be detected. Thus, detection techniques must be sensitive, robust and versatile. Technological advances in NMR spectroscopy [251, 252] and MS have facilitated the use of these two types of techniques to measure the changes in the concentration of small metabolites over time [245, 250, 253-269]. NMR metabolomics methods are inherently quantitative but have a low sensitive and require larger amounts of a metabolite for identification than MS methods. MS is more sensitive but cannot provide absolute quantification. A more detailed analysis of these and other experimental techniques used in metabolomics experiments can be found in the literature [268, 270, 271].

Computational methods are required at every step to process and analyze metabolomics data [263]. A single spectrum in a metabolomics experiment may contain many different signals. This requires powerful computational and statistical methods as well as accurate spectral maps for the many different types of metabolites. Statistical and computational methods are needed to deconvolute the metabolomics spectra. Pattern recognition methods are required when comparing different metabolomics experiments of the same system under different conditions. Methods for pattern recognition are reviewed for example in [248, 249, 272]. The comparisons between different metabolomics profiles help identifying metabolites whose levels are significantly modified, thus pointing at direct changes in regulation of metabolism between the two conditions. Algorithms such as those developed by Vance *et al.* [273] can be applied in determining the causal sequence of events in metabolomics experiments. Parameter fitting methods and algorithms can use the data to obtain parameter values for mathematical models [274-278].

S.7 Interaction Data

Available Data

Interaction data can be divided into three categories:

Genetic and functional interactions. By genetic interactions data we mean data that provide information regarding essentiality of genes (e.g. [279]) or complementation between different genes [280]. For example, in *S. cerevisiae*, a multi-copy vector containing the ISA2 gene rescues the Δgrx5 mutant phenotype [281]. By functional interaction data we mean data that provide information about which processes and functions are genes and proteins involved in. Genetic and functional interaction data provides very useful information regarding the function of the different genes. It may be invaluable in providing information regarding the pathways and responses in which genes of unknown function are involved. ***Prophecy*** is a database that collects this data for *S. cerevisiae* [282]. However, there is, to our knowledge no other resource that systematically collects this information for other organisms. Another type of functional interaction data is that of genes and proteins that are known to belong to specific metabolic pathways, signal transduction pathways and gene circuits. By homology analysis many proteins of any genome can be placed into the different pathways and circuits, thus providing functional information.

Physical interactions. By physical interactions data we mean data that provides information regarding actual physical interactions between different proteins or between proteins and nucleic acid. There are many different types of experiments that provide data regarding these interactions. Automated and large scale Two Hybrid Screens (THS) systematically take the proteins of a genome and scan for physical interactions between each pair of proteins [283-285]. If

an interaction exists, a chemical product that can be detected is produced or photons are emitted [285, 286]. A systematic application of the THS technology has been reported for yeast, worm, fly, human, *H. pylori* and other organisms (see [286-296]). Different reporter methods will provide different sensitivity and robustness in THS [291]. The THS assays also have problems with false positive and false negative interactions. This is so because the interaction between proteins is often measured in conditions that may not be the native conditions under which the two proteins meet in the cell. An alternative method for HTP assays of protein interactions is by using classical biochemical co-purification methods followed by MS identification of the purified proteins [287, 297]. So far there is a 10-20% overlap between dataset of protein interactions that are determined independently [298]. It is conceivable that in the near future protein chips, where the individual proteins of the genome are secured to some support material, will be devised and used to fish out interacting partners from whole cell protein extracts. There are already affinity chips that are used for purification of proteins [211]. Additional sources for protein-protein interaction information are *in silico* predictions. By using sequence information and evolutionary analysis there are methods that predict, for example, co-evolution of residues in different proteins or gene fusion events in different genomes. Predictions of co-evolution or detection of gene fusion events are interpreted as indicating that there is a physical interaction between two proteins (see [299-302] for a review). On-line resources where information from HTP interaction experiments is deposited include the ***BIND*** database [303] the ***PRIME*** database [304], the ***MIPS*** database [305], the ***DIP*** database [306] and the ***MINT*** database [307]. All these databases also contain information regarding small scale interaction experiments, mined from the literature. Some databases focus on specific organisms, as is the case for example of the *E. coli* database ***EchoBase*** [308].

Another type of physical interaction between cellular components that has implications in cell function is the one occurring between proteins and nucleic acids. Information regarding which regulatory motifs are bound by the different transcription factors (TF) in a genome is important because it allows the identification of which genes are regulated by each transcription factor. The development of ChIP-chip technology [309, 310] enables HTP research about which genes are regulated by which TFs in a genome. In ChIP-chip experiments, cells are grown under the conditions of interest and then they are lysed. Then one of a series of methods is used to crosslink DNA and protein, followed by fragmentation of the genomic DNA. The DNA fragments are then co-purified with the transcription factors of interest, amplified, and hybridized with a genomic DNA microarray. This hybridization reveals where in the genome the TF binds, allowing the inference of gene circuits controlled by each TF. As is the case for protein-protein interactions, there is no large scale central repository for this type of information. There are however smaller databases that contain information regarding the different aspects of protein-DNA interaction. The ***NDB***, ***PSIBASE*** and ***AANT*** databases contain information obtained from DNA protein complexes deposited at the PDB [311-313]. ***NPInter*** also has information about interactions between protein and non coding RNA in *E. coli, S. cerevisiae, C. elegans, D. melanogaster, M. musculus,* and *H. sapiens* [314]. The ***BDTNP*** database at Berkeley (https://bdtnp.lbl.gov/Chipper-/index.jsp) is a secured access resource for the analysis of fly related DNA-protein interactions.

Analytical Methods and Tools

The computational methods used for mining interaction data range from those used to find co-occurrence of genes in the literature, to those used to create protein alignments and to calculate co-evolution of specific residues, and to those used for *in silico* protein docking. Many algorithms use graph theory to derive networks of functional interactions from the different types of data. From a mathematical modeler's perspective, the representation of networks obtained from mining interaction data is in general far from perfect. Most representations of these networks show a set of nodes (the genes/proteins) connected by edges whenever there is some form of interaction between them (e.g. [315]). This kind of representation is ambiguous and prevents direct utilization of the networks for the generation of mathematical models. Some tools (e.g. pathway tools [96, 100]) represent networks with diagrams that can be easily and unambiguously parsed as models. However, this is mostly for enzyme network data. Building something similar for systematic reconstruction of gene circuits and signal transduction pathways is difficult until a standardized classification is accepted for signal transduction proteins and transcription factors.

S.8 Kinetic Data

Kinetic data are fundamental for the development of mathematical models in molecular biology. Having such data in abundance and devising good classification systems for the different functions of cellular species facilitates automated model creation. This job has so far been easier for enzymes than for other types of molecules. Many years ago, the enzyme commission (EC) developed a functional classification scheme for the activity of enzymes that, although not perfect, is very useful when annotating proteins with enzyme functions [316]. Databases such as ***BRENDA*** [19, 317-319] include information about the function of different types of enzymes, and in some cases also about kinetic parameters and thermodynamic energy measurements. Kinetic information can also be found for example in ***KDBI*** [320]. ***EzCatDB*** [321, 322] and ***MaCie*** [323, 324] include information regarding catalytic mechanisms. ***ProTherm*** provides thermodynamics data for mutant proteins [325]. However, the kinetic information is not classified and organized in such a way that researchers can directly take the parameter values and use them to build models. Anyone who is interested in doing so must consult the original sources where the parameter values are reported and make sure that units are consistent and that the experimental conditions during the determination are appropriate for the purposes of the model.

It would be important to have similar functional classifications for other types of proteins, such as receptors, structural proteins, transcription factors and so on. This would facilitate automated building of schema that could then be used to build mathematical models. However, proposed standard for such classifications are still far from perfect or universal. Some classifications based upon the DNA binding

motifs or the structural binding domain of transcription factors have been proposed (see e.g. [326, 327]). Classifications similar to that proposed by the EC for enzymes have been proposed for signal transduction molecules [328] and transport proteins [329]. However, their use is far from widespread and it is unclear if they will be adopted as a standard.

It would also be extremely useful to have standard functional classifications for:

a) **Lipid components**. The chemical structure of these lipids may constitute a good basis for such a classification.

b) **Nucleic acid molecules**, including the different types of small RNAs that are currently being shown to play important roles in regulating and catalyzing functional aspects of the cell metabolism. Although a general classification has been proposed for RNA [330], it focuses on structure, rather than function.

c) **Small metabolites**. Again, the chemical structure of these small metabolites may constitute a good basis for a functional classification. The *Ligand* database [331] is a good repository of small metabolic molecules. Other methods, based on chemical and structural information allow for the development of different types of metabolite classifications (e. g. [332]).

Having a classification for the different proteins, genes and other components of the cell has allowed the creation of many resources with more systemic functional information such as lists of pathways, gene circuits and signal transduction pathways (e.g. ***Metacyc*** [95], ***KEGG*** [80, 82] or ***EMP*** [333]). Such resources provide annotated information regarding the function of the different cellular components.

S.9 Mathematical Models

Available Data

Mathematical models of specific molecular processes in different organisms are an increasingly important source of information for the understanding of the systemic behavior in molecular biology. Such models have been published since the 1950s and they are mostly scattered in the literature. Datasets that collect and organize these models have been recently started, but relatively few models are included [334-341]. Model annotation is much more challenging than annotation of genes and proteins for several reasons. For any given system, one can create models using different formalisms and at different levels of detail. This depends, among other things, upon the questions one wants to ask of the model. Models of similar processes on different organisms may involve different enzymes, different number of steps and branch points, etc. Finally, the existence of specific regulatory signals that may vary from one condition to another, from one protein isoform to another, or from one organism to the next, introduces further complexity into the modeling process.

Despite all these difficulties, it is clear that a standard model reporting language is useful to reutilize models or part of models that have been previously defined and curated. Building models is a challenging task and any tools that facilitate the model building process are useful. Currently there is an effort to make newly published models comply with minimal information criteria for publication [338]. There is also a drive to facilitate the interchange of model files between different programs and researchers by asking that models be deposited in central databases, using a standardized modeling language such as ***SBML*** or ***CellML*** [342-345]. A large investment is also being made in the development of tools for model set up and analysis [346-352]. The model database efforts are currently in their infancy and the databases are still small. However, if one is willing to search the literature, one will find many models that are useful in understanding the systemic behavior of molecular biology processes.

Analytical Methods and Tools

There are different formalisms for writing models and different classes of models. The choice of a given mathematical representation depends on the goals of the modeling exercise, the complexity of the target system, and the available data (either in terms of parameters or in terms of time course data, metabolite levels, enzyme activities, etc.). Furthermore, the same problem can be studied at different levels of complexity. For example, one can find models that range from the modeling of a single enzyme reaction to those that attempt to model the whole cell, going through intermediate scale models that consider small sets of cellular pathways and circuits.

Among the smaller scale models that consider only a very limited number of molecules and reactions we find many different types of modeling. Mainly, quantum modeling is used for investigating the active state of an enzyme reaction [353], molecular dynamics is used to study protein folding [354], stochastic modeling can be used for modeling networks of reactions with a small number of molecules [346]. For large scale models of many pathways, the quantum modeling, molecular dynamics modeling, Monte Carlo modeling and stochastic modeling become unfeasible due to lack of computer power. Sets of ordinary and partial differential equations are common in exploring the dynamic behavior of complex pathways [275, 355, 356]. Qualitative types of modeling such as discrete binary mode modeling [357-359] or flux balance analysis [360-363] are also used to study large networks.

Integration of models from different origins and using different mathematical forms is a very challenging problem. Before discussing possible solutions to this problem, it is worth it briefly detail some of the specific requirements of the most common formalisms. A simple description of a process is appealing because one require less information to build a model from such a description. However, the utility of the resulting models may be limited if our goal is to understand dynamic behavior and design principles. More complex descriptions of a process have better changes of producing models that reproduce a wide range of behavior for the system one is studying. However, the complexity of such models combined with lack of information regarding the process may prevent the creation of more detailed models.

A similar argument can be made for mathematical representations. Simple linear representations will facilitate the analysis of models but will prevent the model from capturing non linear behavior that is typical of biological systems. Approximation theory provides many alternative formalisms that combine simplicity with non linearity [364]. We now briefly discuss different types of models used in molecular biology.

Finite State Models

Finite state models analyze the dynamic behavior of networks in which the nodes are allowed only a finite number of states (e.g. on-off). Such a mathematical description is very simple, thus allowing the analysis of complex and large networks. The basic information required to built such models include: (i) the stoichiometric matrix of the reactions taking place in the model, which can be derived from the conceptual scheme of the target system, (ii) the relevant regulatory signals, so that we can appropriately describe the on-off states of the system's elements, and (iii) a set of functions, usually sigmoid, that are used to decide the on-off state of the different elements (e.g. [365]). Finite state models are appealing because they need a minimum of information as input and provide clues about the dynamic properties of a system [358]. However, in general, these models will not be useful if detailed quantitative predictions are needed.

Models Based on Stoichiometry

Flux-balance analysis (FBA) and graph theoretical approaches study network topology and relate changes in that topology to qualitative changes in the dynamics of the system. These models often study the entire metabolic network of an organism [360, 366-381]. As in the case of the finite state models FBA requires a minimum of information, besides the stoichiometric matrix of the model. From this matrix and using graph theoretical methods and optimization techniques, FBA can, for example evaluate the effect of deleting a gene on the flux distribution, thus identifying the expected phenotypes for mutants in the different enzymes. A disadvantage of these approaches is that they fail to account for dynamic regulation. FBA models are a good choice for genome-wide descriptions and can be used as a base for more detailed modeling of specific pathways.

Stochastic Modeling

Stochastic models consider the details of the system and either a) use a Master equation to study the time distribution of the metabolite concentration [382-385], or b) consider every molecule in the system and follow their time evolution using chemical kinetics and algorithms that are able to statistically predict each elementary reaction event in the system. *Gillespie like algorithms* require that all reactions are mass action in order for the methods to work [346-349]. There are now newer stochastic methods that are being developed and loosen this constraint [386]. These models are computationally very heavy. Whenever dealing with processes that involve large pools of molecules, deterministic approaches are good approximations to the more detailed stochastic approach.

Continuous Modeling

The use of differential equations, either ordinary (for well stirred systems) or partial (for systems with spatial differentiation) to define mathematical models that can be used to quantitatively simulate the dynamics of a molecular system was the first [387, 388] and is still the more prevalent choice as a computational modeling tool. There are mathematical rational expressions that simplify complex mechanisms and are considered to represent more or less accurately the dynamics of many different types of individual molecular biology processes. For example, for a simple one substrate enzyme, the Michaelis-Menten rate expression can represent the dynamics of the reaction in well stirred systems [389]. Many other rational expressions have been derived for more complex enzyme atic mechanisms. However, we lack the knowledge of how many processes happen, which prevents the use of these pre-defined kinetic expressions in modeling. Even if the mechanism is known, in most cases the parameter values have not been measured and thus the mathematical representation using rational expression cannot be parameterized. Furthermore, when parameter values are available, the experimental conditions under which they have been measured may invalidate their use.Additional complications occur when one is interested in modeling spatially non-homogenous systems. In such a case PDEs must be used and many of the simplifying assumptions for obtaining rational rate expressions such as the Michaelis Menten kinetics often break down. Thus, depending on the level of mechanistic, spatial and mathematical detail one wants to consider, a given conceptual scheme can generate many different models.

One way to side-step the lack of knowledge regarding mechanism and, in some cases, parameter values is by using approximation theory. Mathematical theory allows us to approximate functions of known and unknown form with structured, canonical, representations that are precise at the operating point of the approximation and accurate over a varying range of values about that operating point. The most widespread and successful approaches use Taylor series to approximate the kinetic functions. This strategy was used by Savageau in the late sixties to generate a non linear representation of molecular biology systems, known as the *Power-law* formalism [390-392]. The Power-law representation can be derived from the conceptual model and some systemic properties can be analyzed without the need of a detailed kinetic characterization [393-406]. Furthermore, we can easily modify a given model by adding terms to any of the equations, which can be an advantage for sharing models. Alternative representations to the Power-law formalism that share some of its advantages are the *Lin-log* [265, 407, 408] and *(log)linear* [409] formalisms. Both formalisms require the same information as required for the Power-law formalism, although the final representation differs [364, 410]. Recently, a *Saturable and Cooperative (SC)* formalism has been derived by using Taylor series approximation after a power-inverse transformation [364, 410]. In a different strategy, the use of a special rate-law called *Convenience kinetics* has been proposed as a way of obtaining a general representation [411]. Because power-law models can be easily and automatically obtained from other representations, the use of this

(and other) formalism(s where such usage is possible) may facilitate HTP modelling building and analysis.

REFERENCES

[1] Richards DE. Medline at thirty-five. *Evid Based Dent* **2006**; 7: 89.

[2] Reid T, Dunikowski LG. PubMed Central--at last. *Can Fam Physician* **2006**; 52: 159-160, 165.

[3] Abulaish M, Dey L. Knowledge enhancement through ontology-guided text mining. *Pattern Recognition and Machine Intelligence, Proceedings* **2005**; 3776: 601-604.

[4] Sevinc A. Web of science: a unique method of cited reference searching. *J Natl Med Assoc* **2004**; 96: 980-983.

[5] Giglia E. Beyond PubMed. Other free-access biomedical databases. *Eura Medicophys* **2007**; 43: 563-569.

[6] De Groote SL. PubMed, Internet Grateful Med, and Ovid: a comparison of three MEDLINE Internet interfaces. *Med Ref Serv Q* **2000**; 19: 1-13.

[7] Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* **2008**; 22: 338-342.

[8] Gall C, Brahmi FA. Retrieval comparison of EndNote to search MEDLINE (Ovid and PubMed) versus searching them directly. *Med Ref Serv Q* **2004**; 23: 25-32.

[9] Janke RG. Current Contents Connect and PubMed--a comparison of content and currency. *Health Info Libr J* **2002**; 19: 230-232.

[10] Sutton VR, Hauser SE. Preliminary comparison of the Essie and PubMed search engines for answering clinical questions using MD on Tap, a PDA-based program for accessing biomedical literature. *AMIA Annu Symp Proc* **2005**: 1128.

[11] Vanhecke TE, Barnes MA, Zimmerman J, Shoichet S. PubMed vs. HighWire Press: a head-to-head comparison of two medical literature search engines. *Comput Biol Med* **2007**; 37: 1252-1258.

[12] Peng Y, Zhang X. Integrative data mining in systems biology: from text to network mining. *Artif Intell Med* **2007**; 41: 83-86.

[13] Ananiadou S, Kell DB, Tsujii J. Text mining and its potential applications in systems biology. *Trends Biotechnol* **2006**; 24: 571-579.

[14] Liu Y, Navathe SB, Civera J, Dasigi V, Ram A, Ciliax BJ, Dingledine R. Text mining biomedical literature for discovering gene-to-gene relationships: a comparative study of algorithms. *IEEE/ACM Trans Comput Biol Bioinform* **2005**; 2: 62-76.

[15] Liu Y, Brandon M, Navathe S, Dingledine R, Ciliax BJ. Text mining functional keywords associated with genes. *Medinfo* **2004**; 11: 292-296.

[16] Natarajan J, Berrar D, Dubitzky W, *et al*. Text mining of full-text journal articles combined with gene expression analysis reveals a relationship between sphingosine-1-phosphate and invasiveness of a glioblastoma cell line. *BMC Bioinformatics* **2006**; 7: 373.

[17] Barthelmes J, Ebeling C, Chang A, Schomburg I, Schomburg D. BRENDA, AMENDA and FRENDA. the enzyme information system in 2007. *Nucleic Acids Res* **2007**; 35: D511-514.

[18] Pharkya P, Nikolaev EV, Maranas CD. Review of the BRENDA Database. *Metab Eng* **2003**; 5: 71-73.

[19] Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* **2004**; 32: D431-433.

[20] Adeva JJG, Calvo R. Mining text with Pimiento. *IEEE Internet Computing* **2006**; 10: 27-35.

[21] Antonakaki D, Kanterakis A, Potamias G. Biomedical literature mining for text classification and construction of gene networks. *Advances in Artificial Intelligence, Proceedings* **2006**; 3955: 469-473.

[22] Choi YJ, Park SS. Efficient classification method for complex biological literature using text and data mining combination. *Intelligent Data Engineering and Automated Learning - Ideal 2006; Proceedings* **2006**; 4224: 688-696.

[23] Davis N, Demetriou G, Gaizauskas R, Guo YK, Roberts I. Web service architectures for text mining: An exploration of the issues via an e-science demonstrator. *Int J Web Ser Res* **2006**; 3: 95-112.

[24] Denicia-Carral C, Montes-y-Gomez M, Villasenor-Pineda L, Hernandez RG. **A** text mining approach for definition question answering. *Adv Natl Lang Proc Proc* **2006**; 4139: 76-86.

[25] Doms A, Jakoniene V, Lambrix P, Schroeder M, Wachter T. Ontologies and text mining as a basis for a semantic web for the life sciences. *Reasoning Web* **2006**; 4126: 164-183.

[26] Erhardt RAA, Schneider R, Blaschke C. Status of text-mining techniques applied to biomedical text. *Drug Discov Today* **2006**; 11: 315-325.

[27] Fan WG, Wallace L, Rich S, Zhang ZJ. Tapping the power of text mining. *Commun Acm* **2006**; 49: 77-82.

[28] Hong K, Park J, Yang J, Park S. Biological text mining for extraction of proteins and their interactions. *Pattern Recognition and Machine Intelligence, Proceedings* **2005**; 3776: 645-648.

[29] Hu XH, Zhang XD, Wu D, Zhou XH, Rumm P. Integration of instance-based learning and text mining for identification of potential virus/bacterium as bio-terrorism weapons. *Intelligence and Security Informatics, Proceedings* **2006**; 3975: 548-553.

[30] Jang H, Song SK, Myaeng SH. Text mining for medical documents using a Hidden Markov Model. *Info Retri Technol Proc* **2006**; 4182: 553-559.

[31] Kim HJ. On text mining algorithms for automated maintenance of hierarchical knowledge directory. *Knowledge Science, Engineering and Management* **2006**; 4092: 202-214.

[32] Kuhn T, Royer L, Fuchs NE, Schroder M. Improving text mining with controlled natural language: A case study for protein interactions. *Data Integration in the Life Sciences, Proceedings* **2006**; **4075**: 66-81.

[33] Lai KK, Yu L, Wang SY. Multi-agent web text mining on the grid for enterprise decision support. *Advanced Web and Network Technologies, and Applications, Proceedings* **2006**; 3842: 540-544.

[34] Liu Y, Navathe SB, Civera J, Dasigi V, Ram A, Ciliax BJ, Dingledine R. Text mining biomedical literature for discovering gene-to-gene relationships: A comparative study of algorithms. *Ieee-Acm Trans Comput Biol Bioinfo* **2005**; 2: 62-76.

[35] Malik R, Franke L, Siebes A. Combination of text-mining algorithms increases the performance. *Bioinformatics* **2006**; 22: 2151-2157.

[36] Rice SB, Nenadic G, Stapley BJ. Mining protein function from text using term-based support vector machines. *BMC Bioinformatics* **2005**; 6 Suppl 1: S22.

[37] SanJuan E, Ibekwe-SanJuan F. Text mining without document context. *Info Proc Manag* **2006**; 42: 1532-1552.

[38] Shaban K, Basir O, Kamel M. Document mining based on semantic understanding of text. *Progress in Pattern Recognition, Image Analysis and Applications, Proceedings* **2006**; 4225: 834-843.

[39] Karamanis N, Lewin I, Seal R, Drysdale R, Briscoe E. Integrating natural language processing with FlyBase curation. *Pac Symp Biocomput* **2007**; 245-256.

[40] Hoffmann R, Valencia A. Implementing the iHOP concept for navigation of biomedical literature. *Bioinformatics* **2005**; 2: ii252-ii258.

[41] Fernandez JM, Hoffmann R, Valencia A. iHOP web services. *Nucleic Acids Res* **2007**; 35: W21-26.

[42] Hoffmann R, Krallinger M, Andres E, Tamames J, Blaschke C, Valencia A. Text mining for metabolic pathways, signaling cascades, and protein networks. *Sci STKE* **2005**; 2005: pe21.

[43] Hoffmann R, Valencia A. Life cycles of successful genes. *Trends Genet* **2003**; 19: 79-81.

[44] Hoffmann R, Valencia A. A gene network for navigating the literature. *Nat Genet* **2004**; 36: 664.

[45] Scherf M, Epple A, Werner T. The next generation of literature analysis: integration of genomic analysis into text mining. *Brief Bioinform* **2005**; 6: 287-297.

[46] Skusa A, Ruegg A, Kohler J. Extraction of biological interaction networks from scientific literature. *Brief Bioinform* **2005**; 6: 263-276.

[47] Stapley BJ, Benoit G. Biobibliometrics: information retrieval and visualization from co-occurrences of gene names in Medline abstracts. *Pac Symp Biocomput* **2000**; 529-540.

[48] Stapley BJ, Kelley LA, Sternberg MJ. Predicting the sub-cellular location of proteins from text using support vector machines. *Pac Symp Biocomput* **2002**; 374-385.

[49] Rogers A, Antoshechkin I, Bieri T, Blasiar D, *et al*. WormBase 2007. *Nucleic Acids Res* **2008**; 36: D612-617.

[50] Crosby MA, Goodman JL, Strelets VB, Zhang P, Gelbart WM. FlyBase: genomes by the dozen. *Nucleic Acids Res* **2007**; 35: D486-491.

[51] Wilson RJ, Goodman JL, Strelets VB. FlyBase: integration and improvements to query tools. *Nucleic Acids Res* **2008**, 36: D588-593.

[52] Chiromatzo AO, Oliveira TY, Pereira G, *et al*. miRNApath: a database of miRNAs, target genes and metabolic pathways. *Genet Mol Res* **2007**; 6: 859-865.
[53] Putz J, Dupuis B, Sissler M, Florentz C. Mamit-tRNA, a database of mammalian mitochondrial tRNA primary and secondary structures. *RNA* **2007**; 13: 1184-1190.
[54] Yin Y, Zhao Y, Wang J, *et al*. antiCODE. a natural sense-antisense transcripts database. *BMC Bioinformatics* **2007**; 8: 319.
[55] Griffiths-Jones S. miRBase: the microRNA sequence database. *Methods Mol Biol* **2006**; 342: 129-138.
[56] Kazakov AE, Cipriano MJ, Novichkov PS, *et al.* RegTransBase--a database of regulatory sequences and interactions in a wide range of prokaryotic genomes. *Nucleic Acids Res* **2007**; 35: D407-412.
[57] Ankeny RA. Sequencing the genome from nematode to human: changing methods, changing science. *Endeavour* **2003**; 27: 87-92.
[58] Gustafson AM, Snitkin ES, Parker SC, DeLisi C, Kasif S. Towards the identification of essential genes using targeted genome sequencing and comparative analysis. *BMC Genomics* **2006**; 7: 265.
[59] Margulies M, Egholm M, Altman WE, *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **2005**; 437: 376-380.
[60] Service RF. Gene sequencing. The race for the $1000 genome. *Science* **2006**; 311: 1544-1546.
[61] Venter JC, Smith HO, Hood L. A new strategy for genome sequencing. *Nature* **1996**; 381: 364-366.
[62] McGinnis S, Madden TL. BLAST. at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **2004**; 32: W20-25.
[63] Kuroda M, Hiramatsu K. Genome sequencing and annotation: an overview. *Methods Mol Biol* **2004**; 266: 29-45.
[64] Besemer J, Borodovsky M. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* **2005**; 33: W451-454.
[65] Cawley SL, Pachter L. HMM sampling and applications to gene finding and alternative splicing. *Bioinformatics* **2003**; 19 Suppl 2: II36-II41.
[66] Chatterji S, Pachter L. Large multiple organism gene finding by collapsed Gibbs sampling. *J Comput Biol* **2005**; 12: 599-608.
[67] Eyras E, Reymond A, Castelo R, Bye JM, *et al.* Gene finding in the chicken genome. *BMC Bioinformatics* **2005**; 6: 131.
[68] Korf I. Gene finding in novel genomes. *BMC Bioinformatics* **2004**; 5: 59.
[69] Noguchi H, Park J, Takagi T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequence**s**. *Nucleic Acids Res* **2006**; 34: 5623-5630.
[70] Wang J, Kraemer E. GFPE. gene-finding program evaluation. *Bioinformatics* **2003**; 19: 1712-1713.
[71] Cuff JA, Coates GM, Cutts TJ, Rae M. The Ensembl computing architecture. *Genome Res* **2004**; 14: 971-975.
[72] Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R, Wittwer P. Metagenomic gene discovery: past, present and future. *Trends Biotechnol* **2005**; 23: 321-329.
[73] Meyer QC, Burton SG, Cowan DA. Subtractive hybridization magnetic bead capture: a new technique for the recovery of full-length ORFs from the metagenome. *Biotechnol J* **2007**; 2: 36-40.
[74] Remington KA, Heidelberg K, Venter JC. Taking metagenomic studies in context. *Trends Microbiol* **2005**; 13: 404.
[75] Rusch DB, Halpern AL, Sutton GK, *et al.* The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **2007**; 5: e77.
[76] Yooseph S, Sutton G, Rusch DB, *et al.* The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol* **2007**; 5: e16.
[77] Hulo N, Bairoch A, Bulliard V. The PROSITE database. *Nucleic Acids Res* **2006**; 34: D227-230.
[78] Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P. Prediction of effective genome size in metagenomic samples. *Genome Biol* **2007**; 8: R10.
[79] Schneider M, Tognolli M, Bairoch A. The Swiss-Prot protein knowledgebase and ExPASy: providing the plant community with high quality proteomic data and tools. *Plant Physiol Biochem* **2004**; 42: 1013-1021.
[80] Kanehisa M. The KEGG database. *Novartis Found Symp* **2002**; 247: 91-101; discussion 101-103, 119-128, 244-152.
[81] Kanehisa M, Goto S, Hattori M, *et al*. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **2006**; 34: D354-357.
[82] Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNe**t**. *Nucleic Acids Res* **2002**; 30: 42-46.
[83] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res* **2004**; 32: D277-280.
[84] Ostell JM, Wheelan SJ, Kans JA. The NCBI data model. *Methods Biochem Anal* **2001**; 43: 19-43.
[85] Guda C. pTARGET. a web server for predicting protein subcellular localization. *Nucleic Acids Res* **2006**; 34: W210-213.
[86] Hoglund A, Donnes P, Blum T, Adolph HW, Kohlbacher O. MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics* **2006**; 22: 1158-1165.
[87] Lee K, Kim DW, Na D, Lee KH, Lee D. PLPD. reliable protein localization prediction from imbalanced and overlapped datasets. *Nucleic Acids Res* **2006**; 34: 4655-4666.
[88] Petsalaki EI, Bagos PG, Litou ZI, Hamodrakas SJ. PredSL. a tool for the N-terminal sequence-based prediction of protein subcellular localization. *Genom Prot Bioinfo* **2006**; 4: 48-55.
[89] Altermann E, Klaenhammer TR. PathwayVoyager: pathway mapping using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. *BMC Genomics* **2005**; 6: 60.
[90] Arakawa K, Kono N, Yamada Y, Mori H, Tomita M. KEGG-based pathway visualization tool for complex omics data. *In Silico Biol* **2005**; 5: 419-423.
[91] Cornwell P. Designing tools for pathway implementation and evaluation. *J Burn Care Rehabil* **1995**; 16: 202-208.
[92] Ekins S, Nikolsky Y, Bugrim A, Kirillov E, Nikolskaya T. Pathway mapping tools for analysis of high content data. *Methods Mol Biol* **2007**; 356: 319-350.
[93] Holford M, Li N, Nadkarni P, Zhao H. VitaPad: visualization tools for the analysis of pathway data. *Bioinformatics* **2005**; 21: 1596-1602.
[94] Kacser H. Kinetic models of development and heredity. *Symp Soc Exp Biol* **1960**; 14: 13-27.
[95] Karp PD, Ouzounis CA, Moore-Kochlacs C, *et al.* Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* 2005; 33: 6083-6089.
[96] Karp PD, Paley S, Romero P. The Pathway Tools software. *Bioinformatics* **2002**; 18 Suppl 1: S225-232.
[97] Klukas C, Schreiber F. Dynamic exploration and editing of KEGG pathway diagrams. *Bioinformatics* **2006.**
[98] Krummenacker M, Paley S, Mueller L, Yan T, Karp PD. Querying and computing with BioCyc databases. *Bioinformatics* **2005**; 21: 3454-3455.
[99] Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **2005**; 21: 3787-3793.
[100] Paley SM, Karp PD. The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Res* **2006**; 34: 3771-3778.
[101] Meier-Kolthoff JP, Auch AF, Huson DH, Goker M. COPYCAT. cophylogenetic analysis tool. *Bioinformatics* **2007**; 23: 898-900.
[102] Berezikov E, Guryev V, Cuppen E. Exploring conservation of transcription factor binding sites with CONREAL. *Methods Mol Biol* **2007**; 395: 437-448.
[103] Brown DP, Krishnamurthy N, Sjolander K. Automated protein subfamily identification and classification. *PLoS Comput Biol* **2007**; 3: e160.
[104] Glanville JG, Kirshner D, Krishnamurthy N, Sjolander K. Berkeley Phylogenomics Group web servers: resources for structural phylogenomic analysis. *Nucleic Acids Res* **2007**; 35: W27-32.
[105] Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res* **2005**; 33: W557-559.
[106] Loots GG, Ovcharenko I. Mulan: multiple-sequence alignment to predict functional elements in genomic sequences. *Methods Mol Biol* **2007**; 395: 237-254.
[107] Meinel T, Krause A, Luz H, Vingron M, Staub E. The SYSTERS Protein Family Database in 2005. *Nucleic Acids Res* **2005**; 33: D226-229.
[108] Tarraga J, Medina I, Arbiza L, *et al*. Phylemon: a suite of web tools for molecular evolution, phylogenetics and phylogenomics. *Nucleic Acids Res* **2007**; 35: W38-42.

[109] Shah JK, Garner HR, White MA, Shames DS, Minna JD. sIR. siRNA Information Resource, a web-based tool for siRNA sequence design and analysis and an open access siRNA database. *BMC Bioinformatics* **2007**; 8: 178.
[110] Abreu-Goodger C, Merino E. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res* **2005**; 33: W690-692.
[111] Bengert P, Dandekar T. Riboswitch finder--a tool for identification of riboswitch RNAs. *Nucleic Acids Res* **2004**; 32: W154-159.
[112] Huang HY, Chien CH, Jen KH, *et al*. an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Res* **2006**; 34: W429-434.
[113] Liu CC, Lin CC, Chen WS, Chen HY, Chang PC, Chen JJ, Yang PC. CRSD. a comprehensive web server for composite regulatory signature discove**ry**. *Nucleic Acids Res* **2006**; 34: W571-577.
[114] Tong J, Jiang P, Lu ZH. RISP. A web-based server for prediction of RNA-binding sites in proteins. *Comput Methods Programs Biomed* **2008**.
[115] Zhang Y. miRU. an automated plant miRNA target prediction server. *Nucleic Acids Res* **2005**; 33: W701-704.
[116] Burley SK, Bonanno JB. Structural genomics. *Methods Biochem Anal* **2003**; 44: 591-612.
[117] Goldsmith-Fischman S, Honig B. Structural genomics: computational methods for structure analysis. *Protein Sci* **2003**; 12: 1813-1821.
[118] Berman HM, Westbrook JD. The impact of structural genomics on the protein data bank. *Am J Pharmacogenomics* **2004**; 4: 247-252.
[119] Berman H, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* **2007**; 35: D301-303.
[120] Bourgeois D, Royant A. Advances in kinetic protein crystallography. *Curr Opin Struct Biol* **2005**; 15: 538-547.
[121] Esposito L, Vitagliano L, Mazzarella L. Recent advances in atomic resolution protein crystallography. *Protein Pept Lett* **2002**; 9: 95-106.
[122] Helliwell JR. Overview and new developments in softer X-ray (2A < lambda < 5A) protein crystallography. *J Synchrotron Radiat* **2004**; 11: 1-3.
[123] Massover WH. Radiation damage to protein specimens from electron beam imaging and diffraction: a mini-review of anti-damage approaches, with special reference to synchrotron X-ray crystallography. *J Synchrotron Radiat* **2007**; 14: 116-127.
[124] Myles DA. Neutron protein crystallography: current status and a brighter future. *Curr Opin Struct Biol* **2006**; 16: 630-637.
[125] Riekel C, Burghammer M, Schertler G. Protein crystallography microdiffraction. *Curr Opin Struct Biol* **2005**; 15: 556-562.
[126] Grishaev A, Llinas M. Protein structure elucidation from minimal NMR data: the CLOUDS approach. *Methods Enzymol* **2005**; 394: 261-295.
[127] Kay LE. NMR studies of protein structure and dynamics. *J Magn Reson* **2005**; 173: 193-207.
[128] Snyder DA, Bhattacharya A, Huang YJ, Montelione GT. Assessing precision and accuracy of protein structures derived from NMR data. *Proteins* **2005**; 59: 655-661.
[129] Takeuchi K, Wagner G. NMR studies of protein interactions. *Curr Opin Struct Biol* **2006**; 16: 109-117.
[130] Wishart D. NMR spectroscopy and protein structure determination: applications to drug discovery and development. *Curr Pharm Biotechnol* **2005**; 6: 105-120.
[131] Burkhardt K, Schneider B, Ory J. A biocurator perspective: annotation at the Research Collaboratory for Structural Bioinformatics Protein Data Bank. *PLoS Comput Biol* **2006**; 2: e99.
[132] Sussman JL, Abola EE, Lin D, Jiang J, Manning NO, Prilusky J. The protein data bank. Bridging the gap between the sequence and 3D structure worl**d**. *Genetica* **1999**; 106: 149-158.
[133] Szabadka Z, Grolmusz V. High throughput processing of the structural information in the protein data bank. *J Mol Graph Model* **2006**.
[134] Bernstein HJ. Recent changes to RasMol, recombining the variants. *Trends Biochem Sci* **2000**; 25: 453-455.
[135] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL. An automated protein homology-modeling server. *Nucleic Acids Res* **2003**; 31: 3381-3385.
[136] Dobson PD, Cai YD, Stapley BJ, Doig AJ. Prediction of protein function in the absence of significant sequence similarity. *Curr Med Chem* **2004**; 11: 2135-2142.
[137] Godzik A. Fold recognition methods. *Methods Biochem Anal* **2003**; 44: 525-546.
[138] Bates PA, Kelley LA, MacCallum RM, Sternberg MJ. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins* **2001**; Suppl 5: 39-46.
[139] Bourne PE. CASP and CAFASP experiments and their findings. *Methods Biochem Anal* **2003**; 44: 501-507.
[140] Falcon H, Crosse A, Donaghy J, *et al.* CASP and CONSORT. *Br Dent J* **2006**; 201: 130; author reply 130-131.
[141] Kryshtafovych A, Venclovas C, Fidelis K, Moult J. Progress over the first decade of CASP experiments. *Proteins* **2005**; 61 Suppl 7: 225-236.
[142] Moult J. A decade of CASP. progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struct Biol* **2005**; 15: 285-289.
[143] Moult J, Fidelis K, Rost B, Hubbard T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)--round 6. *Proteins* **2005**; 61 Suppl 7: 3-7.
[144] Valencia A. Protein refinement: a new challenge for CASP in its 10th anniversary. *Bioinformatics* **2005**; 21: 277.
[145] Venclovas C, Zemla A, Fidelis K, Moult J. Assessment of progress over the CASP experiments. *Proteins* **2003**; 53 Suppl 6: 585-595.
[146] Chivian D, Kim DE, Malmstrom L, Schonbrun J, Rohl CA, Baker D. Prediction of CASP6 structures using automated Robetta protocols. *Proteins* **2005**; 61 Suppl 7: 157-166.
[147] Buzko OV, Bishop AC, Shokat KM. Modified AutoDock for accurate docking of protein kinase inhibitors. *J Comput Aided Mol Des* **2002**; 16: 113-127.
[148] Carter P, Lesk VI, Islam SA, Sternberg MJ. Protein-protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI. *Proteins* **2005**; 60: 281-288.
[149] Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res* **2004**; 32: W96-99.
[150] Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **2004**; 20: 45-50.
[151] Fahmy A, Wagner G. TreeDock: a tool for protein docking based on minimizing van der Waals energies. *J Am Chem Soc* **2002**; 124: 1241-1250.
[152] Gao Y, Douguet D, Tovchigrechko A, Vakser IA. DOCK-GROUND system of databases for protein recognition studies: unbound structures for docking. *Proteins* **2007**; 69: 845-851.
[153] Goto J, Kataoka R, Hirayama N. Ph4Dock: pharmacophore-based protein-ligand docking. *J Med Chem* **2004**; 47: 6804-6811.
[154] Grosdidier A, Zoete V, Michielin O. EADock: docking of small molecules into protein active sites with a multiobjective evolutionary optimization. *Proteins* **2007**; 67: 1010-1025.
[155] Gupta A, Gandhimathi A, Sharma P, Jayaram B. ParDOCK. an all atom energy based Monte Carlo docking protocol for protein-ligand complexes. *Protein Pept Lett* **2007**; 14: 632-646.
[156] Lensink MF, Mendez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins* **2007**; 69: 704-718.
[157] Martin L, Catherinot V, Labesse G. kinDOCK. a tool for comparative docking of protein kinase ligands. *Nucleic Acids Res* **2006**; 34: W325-329.
[158] May A, Zacharias M. Protein-protein docking in CAPRI using ATTRACT to account for global and local flexibility. *Proteins* **2007**; 69: 774-780.
[159] Pierce B, Weng Z. ZRANK. reranking protein docking predictions with an optimized energy function. *Proteins* **2007**; 67: 1078-1086.
[160] Seifert MH, Schmitt F, Herz T, Kramer B. ProPose: a docking engine based on a fully configurable protein-ligand interaction model. *J Mol Model* **2004**; 10: 342-357.
[161] Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res* **2006**; 34: W310-314.
[162] van Dijk M, van Dijk AD, Hsu V, Boelens R, Bonvin AM. Information-driven protein-DNA docking using HADDOCK. it is a matter of flexibility. *Nucleic Acids Res* **2006**; 34: 3317-3325.
[163] Zacharias M. ATTRACT. protein-protein docking in CAPRI using a reduced protein model. *Proteins* **2005**; 60: 252-256.
[164] Ritchie DW. Evaluation of protein docking predictions using Hex 3.1 in CAPRI rounds 1 and 2. *Proteins* **2003**; 52: 98-106.

[165] Barrett T, Troup DB, Wilhite SE, *et al.* NCBI GEO. mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res* **2007**; 35: D760-765.
[166] Kasai Y, Hashimoto S, Yamada T, *et al*. 5' SAGE -end Serial Analysis of Gene Expression database. *Nucleic Acids Res* **2005**; 33: D550-552.
[167] Parkinson H, Kapushesky M, Shojatalab M, *et al.* ArrayExpress--a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res* **2007**; 35: D747-750.
[168] Tateno Y, Ikeo K. [International public gene expression database (CIBEX) and data submission]. *Tanpakushitsu Kakusan Koso* **2004**; 49: 2678-2683.
[169] Ikeo K, Ishi-i J, Tamura T, Gojobori T, Tateno Y. CIBEX. center for information biology gene expression database. *C R Biol* **2003**; 326: 1079-1082.
[170] Rocca-Serra P, Brazma A, Parkinson H, Sarkans U, *et al.* ArrayExpress: a public database of gene expression data at EBI. *C R Biol* **2003**; 326: 1075-1078.
[171] Kodzius R, Matsumura Y, Kasukawa TS, *et al.* Absolute expression values for mouse transcripts: re-annotation of the READ expression database by the use of CAGE and EST sequence tags. *FEBS Lett* **2004**; 559: 22-26.
[172] Xia Y, Campen A, Rigsby D, *et al*. DGEM--a microarray gene expression database for primary human disease tissues. *Mol Diagn Ther* **2007**; 11: 145-149.
[173] Zhang Y, Luoh SM, Hon LS, Baertsch R, Wood WI, Zhang Z. GeneHub-GEPIS. digital expression profiling for normal and cancer tissues based on an integrated gene database. *Nucleic Acids Res* **2007**; 35: W152-158.
[174] Smith CM, Finger JH, Hayamizu TF, *et al*. The mouse Gene Expression Database (GXD): 2007 update. *Nucleic Acids Res* **2007**; 35: D618-623.
[175] Jen CH, Manfield IW, Michalopoulos I, *et al*. The Arabidopsis co-expression tool (ACT): a WWW-based tool and database for microarray-based gene expression analysis. *Plant J* **2006**; 46: 336-348.
[176] Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **1995**; 270: 467-470.
[177] Shalon D, Smith SJ, Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* **1996**; 6: 639-645.
[178] Nguyen DV, Arpat AB, Wang N, Carroll RJ. DNA microarray experiments: biological and technological aspects. *Biometrics* **2002**; 58: 701-717.
[179] Dai JL. Serial analyses of gene expression (SAGE). *Methods Mol Med* **2005**; 103: 161-174.
[180] Knox DP, Skuce PJ. SAGE and the quantitative analysis of gene expression in parasites. *Trends Parasitol* **2005**; 21: 322-326.
[181] Kodzius R, Kojima M, Nishiyori H, *et al.* CAGE. cap analysis of gene expression. *Nat Methods* **2006**; 3: 211-222.
[182] Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **2005**; 85: 1-15.
[183] Tannenbaum E. An RNA-centered view of eukaryotic cells. *Biosystems* **2006**; 84: 217-224.
[184] Bueno Filho JS, Gilmour SG, Rosa GJ. Design of microarray experiments for genetical genomics studies. *Genetics* **2006**; 174: 945-957.
[185] Imbeaud S, Auffray C. 'The 39 steps' in gene expression profiling: critical issues and proposed best practices for microarray experiments. *Drug Discov Today* **2005**; 10: 1175-1182.
[186] Blalock EM, Chen KC, Stromberg AJ, *et al*. Harnessing the power of gene microarrays for the study of brain aging and Alzheimer's disease: statistical reliability and functional correlation. *Ageing Res Rev* **2005**; 4: 481-512.
[187] Leung YF, Cavalieri D. Fundamentals of cDNA microarray data analysis. *Trends Genet* **2003**; 19: 649-659.
[188] Olson NE. The microarray data analysis process: from raw data to biological significance. *NeuroRx* **2006**; 3: 373-383.
[189] Pounds S, Cheng C. Statistical development and evaluation of microarray gene expression data filters. *J Comput Biol* **2005**; 12: 482-495.
[190] Barrett T, Edgar R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods Enzymol* **2006**; 411: 352-369.
[191] Saeed AI, Bhagabati NK, Braisted JC, *et al*. TM4 microarray software suite. *Methods Enzymol* **2006**; 411: 134-193.
[192] Troein C, Vallon-Christersson J, Saal LH. An introduction to BioArray Software Environment. *Methods Enzymol* **2006**; 411: 99-119.
[193] Reimers M, Carey VJ. Bioconductor: an open source framework for bioinformatics and computational biology. *Methods Enzymol* **2006**; 411: 119-134.
[194] Beer I, Barnea E, Admon A. Centralized data analysis of a large interlaboratory proteomics project: a feasibility study. *Proteomics* **2005**; 5: 3491-3496.
[195] Garwood K, Garwood C, Hedeler C, *et al*. Model-driven user interfaces for bioinformatics data resources: regenerating the wheel as an alternative to reinventing it. *BMC Bioinformatics* **2006**; 7: 532.
[196] Garwood K, McLaughlin T, Garwood C, *et al.* PEDRo: a database for storing, searching and disseminating experimental proteomics data. *BMC Genomics* **2004**; 5: 68.
[197] Hill A, Kim H. The UAB Proteomics Database. *Bioinformatics* **2003**; 19: 2149-2151.
[198] Kremer A, Schneider R, Terstappen GC. A bioinformatics perspective on proteomics: data storage, analysis, and integration. *Biosci Rep* **2005**; 25: 95-106.
[199] Martens L, Hermjakob H, Jones P, *et al*. PRIDE. the proteomics identifications database. *Proteomics* **2005**; 5: 3537-3545.
[200] Martens L, Nesvizhskii AI, Hermjakob H, *et al*. Do we want our data raw? Including binary mass spectrometry data in public proteomics data repositories. *Proteomics* **2005**; 5: 3501-3505.
[201] Jones P, Cote RG, Martens L, *et al.* PRIDE. a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res* **2006**; 34: D659-663.
[202] Desiere F, Deutsch EW, King NL, *et al*. The PeptideAtlas project. *Nucleic Acids Res* **2006**; 34: D655-658.
[203] Beavis RC. Using the global proteome machine for protein identification. *Methods Mol Biol* **2006**; 328: 217-228.
[204] Linding R, Jensen LJ, Pasculescu A, *et al*. NetworKIN. a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res* **2008**, 36: D695-699.
[205] Proteomics. *Nat Biotechnol* **2000**; 18 Suppl: IT45-46.
[206] Aggarwal K, Choe LH, Lee KH. Shotgun proteomics using the iTRAQ isobaric tags. *Brief Funct Genomic Proteomic* **2006**; 5: 112-120.
[207] Beynon RJ. The dynamics of the proteome: strategies for measuring protein turnover on a proteome-wide scale. *Brief Funct Genomic Proteomic* **2005**; 3: 382-390.
[208] Beynon RJ, Pratt JM. Strategies for measuring dynamics: the temporal component of proteomics. *Methods Biochem Anal* **2006**; 49: 15-25.
[209] Celis JE, Kruhoffer M, Gromova I, *et al.* Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett* **2000**; 480: 2-16.
[210] DeVoe DL, Lee CS. Microfluidic technologies for MALDI-MS in proteomics. *Electrophoresis* **2006**; 27: 3559-3568.
[211] Elrick MM, Walgren JL, Mitchell MD, Thompson DC. Proteomics: recent applications and new technologies. *Basic Clin Pharmacol Toxicol* **2006**; 98: 432-441.
[212] Gevaert K, Vandekerckhove J. Protein identification methods in proteomics. *Electrophoresis* **2000**; 21: 1145-1154.
[213] Haoudi A, Bensmail H. Bioinformatics and data mining in proteomics. *Expert Rev Proteomics* **2006**; 3: 333-343.
[214] Harry JL, Wilkins MR, Herbert BR, Packer NH, Gooley AA, Williams KL. Proteomics: capacity versus utility. *Electrophoresis* **2000**; 21: 1071-1081.
[215] Hayter JR, Robertson DH, Gaskell SJ, Beynon RJ. Proteome analysis of intact proteins in complex mixtures. *Mol Cell Proteomics* **2003**; 2: 85-95.
[216] Hogan JM, Higdon R, Kolker E. Experimental standards for high-throughput proteomics. *Omics* **2006**; 10: 152-157.
[217] Jensen ON. Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* **2006**; 7: 391-403.
[218] Kellner R. Proteomics. Concepts and perspectives. *Fresenius J Anal Chem* **2000**; 366: 517-524.
[219] Nelson RW, Nedelkov D, Tubbs KA. Biosensor chip mass spectrometry: a chip-based proteomics approach. *Electrophoresis* **2000**; 21: 1155-1163.
[220] Smith JC, Figeys D. Proteomics technology in systems biology. *Mol Biosyst* **2006**; 2: 364-370.

[221] Suter B, Auerbach D, Stagljar I. Yeast-based functional genomics and proteomics technologies: the first 15 years and beyond. *Biotechniques* **2006**; 40: 625-644.
[222] Topaloglou T. Informatics solutions for high-throughput proteomics. *Drug Discov Today* **2006**; 11: 509-516.
[223] Uttamchandani M, Wang J, Yao SQ. Protein and small molecule microarrays: powerful tools for high-throughput proteomics. *Mol Biosyst* **2006**; 2: 58-68.
[224] Wasinger V. Holistic biology of microorganisms: genomics, transcriptomics, and proteomics. *Methods Biochem Anal* **2006**; 49: 3-14.
[225] Wittmann-Liebold B, Graack HR, Pohl T. Two-dimensional gel electrophoresis as tool for proteomics studies in combination with protein identification by mass spectrometry. *Proteomics* **2006**; 6: 4688-4703.
[226] Ram RJ, Verberkmoes NC, Thelen MP, *et al*. Community proteomics of a natural microbial biofilm. *Science* **2005**; 308: 1915-1920.
[227] McCarthy FM, Burgess SC, van den Berg BH, Koter MD, Pharr GT. Differential detergent fractionation for non-electrophoretic eukaryote cell proteomics. *J Proteome Res* **2005**; 4: 316-324.
[228] Vener AV, Stralfors P. Vectorial proteomics. *IUBMB Life* **2005**; 57: 433-440.
[229] Whitelegge JP. Mass spectrometry for high throughput quantitative proteomics in plant research: lessons from thylakoid membranes. *Plant Physiol Biochem* **2004**; 42: 919-927.
[230] Doherty MK, McClean L, Edwards I, *et al.* Protein turnover in chicken skeletal muscle: understanding protein dynamics on a proteome-wide scale. *Br Poult Sci* **2004**; 45 Suppl 1: S27-28.
[231] Doherty MK, Whitehead C, McCormack H, Gaskell SJ, Beynon RJ. Proteome dynamics in complex organisms: using stable isotopes to monitor individual protein turnover rates. *Proteomics* **2005**; 5: 522-533.
[232] Artavanis-Tsakonas K, Misaghi S, Comeaux CA, *et al*. Identification by functional proteomics of a deubiquitinating/deNeddylating enzyme in Plasmodium falciparum. *Mol Microbiol* **2006**; 61: 1187-1195.
[233] Camerini S, Polci ML, Bachi A. Proteomics approaches to study the redox state of cysteine-containing proteins. *Ann Ist Super Sanita* **2005**; 41: 451-457.
[234] Gruhler A, Olsen JV, Mohammed S, *et al.* Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* **2005**; 4: 310-327.
[235] Hjerrild M, Gammeltoft S. Phosphoproteomics toolbox: computational biology, protein chemistry and mass spectrometry. *FEBS Lett* **2006**; 580: 4764-4770.
[236] Julka S, Regnier F. Quantification in proteomics through stable isotope coding: a review. *J Proteome Res* **2004**; 3: 350-363.
[237] Kwon SJ, Choi EY, Choi YJ, Ahn JH, Park OK. Proteomics studies of post-translational modifications in plants. *J Exp Bot* **2006**; 57: 1547-1551.
[238] Nielsen ML, Savitski MM, Zubarev RA. Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. *Mol Cell Proteomics* **2006**; 5: 2384-2391.
[239] Stern DF. Phosphoproteomics. *Exp Mol Pathol* **2001**; 70: 327-331.
[240] Tannu NS, Hemby SE. Chapter 3 Methods for proteomics in neuroscience. *Prog Brain Res* **2006**; 158: 41-82.
[241] Cui Q, Lewis IA, Hegeman AD, *et al.* Metabolite identification via the Madison Metabolomics Consortium Database. *Nat Biotechnol* **2008**, 26: 162-164.
[242] Cottingham K. New database for human metabolomics. *J Proteome Res* **2007**; 6: 1243.
[243] Kotera M, McDonald AG, Boyce S, Tipton KF. Functional group and substructure searching as a tool in metabolomics. *PLoS ONE* 2008, 3: e1537.
[244] Markley JL, Anderson ME, Cui Q, Eghbalnia HR, *et al.* New bioinformatics resources for metabolomics. *Pac Symp Biocomput* **2007**: 157-168.
[245] Castle AL, Fiehn O, Kaddurah-Daouk R, Lindon JC. Metabolomics Standards Workshop and the development of international standards for reporting metabolomics experimental results. *Brief Bioinform* **2006**; 7: 159-165.
[246] Fiehn O, Kristal B, van Ommen B, *et al.* Establishing reporting standards for metabolomic and metabonomic studies: a call for participation. *Omics* **2006**; 10: 158-163.
[247] Jenkins H, Hardy N, Beckmann M, *et al.* A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol* **2004**; 22: 1601-1606.
[248] Lindon JC, Holmes E, Nicholson JK. So what's the deal with metabonomics? *Anal Chem* **2003**; 75: 384A-391A.
[249] Lindon JC, Nicholson JK, Holmes, *et al.* Summary recommendations for standardization and reporting of metabolic analyses. *Nat Biotechnol* **2005**; 23: 833-838.
[250] Kell DB. Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol* **2004**; **7**: 296-307.
[251] Fiehn O. Metabolomics--the link between genotypes and phenotypes. *Plant Mol Biol* **2002**; 48: 155-171.
[252] Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999**; 29: 1181-1189.
[253] Brown JA, Sherlock G, Myers CL, *et al*. Global analysis of gene function in yeast by quantitative phenotypic profiling. *Mol Syst Biol* **2006**; 2: 2006 0001.
[254] Brown SC, Kruppa G, Dasseux JL. Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrom Rev* **2005**; 24: 223-231.
[255] Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol* **2004**; 22: 245-252.
[256] Griffin JL. The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? *Philos Trans R Soc Lond B Biol Sci* **2006**; 361: 147-161.
[257] Grivet JP, Delort AM, Portais JC. NMR and microbiology: from physiology to metabolomics. *Biochimie* **2003**; 85: 823-840.
[258] Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. *Proteomics* **2006**; 6: 4716-4723.
[259] Lee JM, Gianchandani EP, Papin JA. Flux balance analysis in the era of metabolomics. *Brief Bioinform* **2006**; 7: 140-150.
[260] Reo NV. NMR-based metabolomics. *Drug Chem Toxicol* **2002**; 25: 375-382.
[261] Rochfort S. Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research. *J Nat Prod* **2005**; 68: 1813-1820.
[262] Sanford K, Soucaille P, Whited G, Chotani G. Genomics to fluxomics and physiomics - pathway engineering. *Curr Opin Microbiol* **2002**; 5: 318-322.
[263] Shulaev V. Metabolomics technology and bioinformatics. *Brief Bioinform* **2006**; **7**: 128-139.
[264] van der Werf MJ, Jellema RH, Hankemeier T. Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. *J Ind Microbiol Biotechnol* **2005**; 32: 234-252.
[265] Wang L, Birol I, Hatzimanikatis V. Metabolic control analysis under uncertainty: framework development and case studies. *Biophys J* **2004**; 87: 3750-3763.
[266] Wang M, Lamers RJ, Korthout HA, *et al*. Metabolomics in the context of systems biology: bridging traditional Chinese medicine and molecular pharmacology. *Phytother Res* **2005**; 19: 173-182.
[267] Wang QZ, Wu CY, Chen T, Chen X, Zhao XM. Integrating metabolomics into a systems biology framework to exploit metabolic complexity: strategies and applications in microorganisms. *Appl Microbiol Biotechnol* **2006**; 70: 151-161.
[268] Weckwerth W. Metabolomics in systems biology. *Annu Rev Plant Biol* **2003**; 54: 669-689.
[269] Weckwerth W, Morgenthal K. Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today* **2005**; 10: 1551-1558.
[270] Morgenthal K, Weckwerth W, Steuer R. Metabolomic networks in plants: Transitions from pattern recognition to biological interpretation. *Biosystems* **2006**; 83: 108-117.
[271] Morgenthal K, Wienkoop S, Wolschin F, Weckwerth W. Integrative profiling of metabolites and proteins: improving pattern recognition and biomarker selection for systems level approaches. *Methods Mol Biol* **2007**; 358: 57-75.
[272] Holmes E, Antti H. Chemometric contributions to the evolution of metabonomics: mathematical solutions to characterising and interpreting complex biological NMR spectra. *Analyst* **2002**; 127: 1549-1557.
[273] Vance W, Arkin A, Ross J. Determination of causal connectivities of species in reaction networks. *Proc Natl Acad Sci USA* **2002**; 99: 5816-5821.

[274] Marino S, Voit EO. An automated procedure for the extraction of metabolic network information from time series data. *J Bioinform Comput Biol* **2006**; 4: 665-691.
[275] Voit EO, Almeida J, Marino S, *et al*. Regulation of glycolysis in Lactococcus lactis: an unfinished systems biological case study. *Syst Biol (Stevenage)* **2006**; 153: 286-298.
[276] Chou IC, Martens H, Voit EO. Parameter estimation in biochemical systems models with alternating regression. *Theor Biol Med Model* **2006**; 3: 25.
[277] Voit EO, Almeida J. Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics* **2004**; 20: 1670-1681.
[278] Voit EO, Marino S, Lall R. Challenges for the identification of biological systems from *in vivo* time series data. *In Silico Biol* **2005**; 5: 83-92.
[279] Gerdes SY, Scholle MD, Campbell JW, *et al.* Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. *J Bacteriol* **2003**; 185: 5673-5684.
[280] Tong AH, Evangelista M, Parsons AB, *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **2001**; 294: 2364-2368.
[281] Rodriguez-Manzaneque MT, Tamarit J, Belli G, Ros J, Herrero E. Grx5 is a mitochondrial glutaredoxin required for the activity of iron/sulfur enzymes. *Mol Biol Cell* **2002**; 13: 1109-1121.
[282] Fernandez-Ricaud L, Warringer J, Ericson E, *et al*. PROPHECY--a database for high-resolution phenomics. *Nucleic Acids Res* **2005**; 33: D369-373.
[283] Buckholz RG, Simmons CA, Stuart JM, Weiner MP. Automation of yeast two-hybrid screening. *J Mol Microbiol Biotechnol* **1999**; 1: 135-140.
[284] Ding X, Zhang Y, Song WY. Use of rolling-circle amplification for large-scale yeast two-hybrid analyses. *Methods Mol Biol* **2006**; 354: 85-98.
[285] Fields S. High-throughput two-hybrid analysis. The promise and the peril. *FEBS J* **2005**; 272: 5391-5399.
[286] Parrish JR, Gulyas KD, Finley RL, Jr. Yeast two-hybrid contributions to interactome mapping. *Curr Opin Biotechnol* **2006**; 17: 387-393.
[287] Gavin AC, Bosche M, Krause R, *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **2002**; 415: 141-147.
[288] Ito T, Chiba T, Yoshida M. Exploring the protein interactome using comprehensive two-hybrid projects. *Trends Biotechnol* **2001**; 19: S23-27.
[289] Ito T, Ota K, Kubota H, *et al.* Roles for the two-hybrid system in exploration of the yeast protein interactome. *Mol Cell Proteomics* **2002**; 1: 561-566.
[290] Rain JC, Selig L, De Reuse H, *et al.* The protein-protein interaction map of Helicobacter pylori. *Nature* **2001**; 409: 211-215.
[291] Starling AL, Ortega JM, Gollob KJ, *et al.* Evaluation of alternative reporter genes for the yeast two-hybrid system. *Genet Mol Res* **2003**; 2: 124-135.
[292] Stelzl U, Worm U, Lalowski M, *et al*. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **2005**; 122: 957-968.
[293] Uetz P, Hughes RE. Systematic and large-scale two-hybrid screens. *Curr Opin Microbiol* **2000**; 3: 303-308.
[294] Walhout AJ, Boulton SJ, Vidal M. Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. *Yeast* **2000**; **17**: 88-94.
[295] Walhout AJ, Sordella R, Lu X, *et al.* Protein interaction mapping in C. elegans using proteins involved in vulval development. *Science* **2000**; 287: 116-122.
[296] Walhout AJ, Vidal M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* **2001**; 24: 297-306.
[297] Ho Y, Gruhler A, Heilbut A, *et al.* Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **2002**; 415: 180-183.
[298] Santonico E, Castagnoli L, Cesareni G. Methods to reveal domain networks. *Drug Discov Today* **2005**; 10: 1111-1117.
[299] Izarzugaza JM, Juan D, Pons C, Ranea JA, Valencia A, Pazos F. TSEMA. interactive prediction of protein pairings between interacting families. *Nucleic Acids Res* **2006**; 34: W315-319.
[300] Juan D, Pazos F, Valencia A. High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proc Natl Acad Sci USA* **2008,** 105: 934-939.
[301] Pazos F, Rausell A, Valencia A. Phylogeny-independent detection of functional residues. *Bioinformatics* **2006**; 22: 1440-1448.
[302] Pazos F, Valencia A. In silico two-hybrid system for the selection of physically interacting protein pairs. *Proteins* **2002**; 47: 219-227.
[303] Gilbert D. Biomolecular interaction network database. *Brief Bioinform* **2005**; 6: 194-198.
[304] Koike A, Takagi T. PRIME. automatically extracted PRotein Interactions and Molecular Information databasE. *In Silico Biol* **2005**; 5: 9-20.
[305] Pagel P, Kovac S, Oesterheld M, *et al*. The MIPS mammalian protein-protein interaction database. *Bioinformatics* **2005**; 21: 832-834.
[306] Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* **2002**; 30: 303-305.
[307] Chatr-aryamontri A, Ceol A, Palazzi LM, *et al*. MINT. the Molecular INTeraction database. *Nucleic Acids Res* **2007**; 35: D572-574.
[308] Misra RV, Horler RS, Reindl W, Goryanin, II, Thomas GH. EchoBASE. an integrated post-genomic database for *Escherichia coli*. *Nucleic Acids Res* **2005**; 33: D329-333.
[309] Buck MJ, Lieb JD. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **2004**; 83: 349-360.
[310] Wu J, Smith LT, Plass C, Huang TH. ChIP-chip comes of age for genome-wide functional analysis. *Cancer Res* **2006**; 66: 6899-6902.
[311] Fujii S. Database and structural characterization of intermolecular interactions in nucleic acid and protein complex. *Nucleic Acids Symp Ser (Oxf)* **2004**: 127-128.
[312] Gong S, Yoon G, Jang I, Bolser D, *et al.* PSIbase: a database of Protein Structural Interactome map (PSIMAP). *Bioinformatics* **2005**; 21: 2541-2543.
[313] Hoffman MM, Khrapov MA, Cox JC, Yao J, Tong L, Ellington AD. AANT. the Amino Acid-Nucleotide Interaction Database. *Nucleic Acids Res* **2004**; 32: D174-181.
[314] Wu T, Wang J, Liu C, Zhang Y, *et al.* NPInter: the noncoding RNAs and protein related biomacromolecules interaction database. *Nucleic Acids Res* **2006**; 34: D150-152.
[315] Mak HC, Daly M, Gruebel B, Ideker T. CellCircuits: a database of protein network models. *Nucleic Acids Res* **2007**; 35: D538-545.
[316] [IUPAC-IUB Commission on Biochemical Nomenclature (CBN). Nomenclature of multiple enzyme types. Recommendations 1971]. *Hoppe Seylers Z Physiol Chem* **1972**, 353: 852-854.
[317] Sabatti C, Rohlin L, Lange K, Liao JC. Vocabulon: a dictionary model approach for reconstruction and localization of transcription factor binding sites. *Bioinformatics* **2005**; 21: 922-931.
[318] Schomburg I, Chang A, Hofmann O, Ebeling C, Ehrentreich F, Schomburg D. BRENDA. a resource for enzyme data and metabolic information. *Trends Biochem Sci* **2002**; **27**: 54-56.
[319] Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res* **2002**; 30: 47-49.
[320] Ji ZL, Chen X, Zhen CJ, *et al.* KDBI. Kinetic Data of Biomolecular Interactions database. *Nucleic Acids Res* **2003**; 31: 255-257.
[321] Nagano N. EzCatDB. the Enzyme Catalytic-mechanism Database. *Nucleic Acids Res* **2005**; 33: D407-412.
[322] Nagano N, Noguchi T, Akiyama Y. Systematic comparison of catalytic mechanisms of hydrolysis and transfer reactions classified in the EzCatDB database. *Proteins* **2007**; 66: 147-159.
[323] Holliday GL, Almonacid DE, Bartlett GJ, O'Boyle N M, Torrance JW, Murray-Rust P, Mitchell JB, Thornton JM. MACiE (Mechanism, Annotation and Classification in Enzymes): novel tools for searching catalytic mechanisms. *Nucleic Acids Res* **2006.**
[324] Holliday GL, Bartlett GJ, Almonacid DE, *et al*. MACiE. a database of enzyme reaction mechanisms. *Bioinformatics* **2005**; 21: 4315-4316.
[325] Kumar MD, Bava KA, Gromiha MM, *et al*. ProTherm and ProNIT. thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res* **2006**; 34: D204-206.
[326] Rigali S, Schlicht M, Hoskisson P, *et al*. Extending the classification of bacterial transcription factors beyond the helix-turn-helix

motif as an alternative approach to discover new cis/trans relationships. *Nucleic Acids Res* **2004**; 32: 3418-3426.
[327] Stegmaier P, Kel AE, Wingender E. Systematic DNA-binding domain classification of transcription factors. *Genome Inform* **2004**; 15: 276-286.
[328] Chen M, Lin S, Hofestaedt R. STCDB. Signal Transduction Classification Database. *Nucleic Acids Res* **2004**; 32: D456-458.
[329] Saier MH, Jr.: A functional-phylogenetic system for the classification of transport proteins. *J Cell Biochem* **1999**; Suppl 32-33: 84-94.
[330] Tamura M, Hendrix DK, Klosterman PS, Schimmelman NR, Brenner SE, Holbrook SR. SCOR. Structural Classification of RNA, version 2.0. *Nucleic Acids Res* **2004**; 32: D182-184.
[331] von Grotthuss M, Koczyk G, Pas J, Wyrwicz LS, Rychlewski L. Ligand.Info small-molecule Meta-Database. *Comb Chem High Throughput Screen* **2004**; 7: 757-761.
[332] Ceroni A, Costa F, Frasconi P. Classification of small molecules by two- and three-dimensional decomposition kernels. *Bioinformatics* **2007**; 23: 2038-2045.
[333] Selkov E, Basmanova S, Gaasterland T, *et al*. The metabolic pathway collection from EMP. the enzymes and metabolic pathways database. *Nucleic Acids Res* **1996**; 24: 26-28.
[334] Campagne F, Neves S, Chang CW, Skrabanek L, Ram PT, Iyengar R, Weinstein H. Quantitative information management for the biochemical computation of cellular networks. *Sci STKE* **2004**; pl11.
[335] Keseler IM, Collado-Vides J, Gama-Castro S, *et al*. EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Res* **2005**; 33: D334-337.
[336] Le Novere N. Model storage, exchange and integration. *BMC Neurosci* **2006**; 7 Suppl 1: S11.
[337] Le Novere N, Bornstein B, Broicher A, *et al*. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res* **2006**; 34: D689-691.
[338] Le Novere N, Finney A, Hucka M, *et al*. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* **2005**; 23: 1509-1515.
[339] Olivier BG, Snoep JL. Web-based kinetic modelling using JWS Online. *Bioinformatics* **2004**; 20: 2143-2144.
[340] Salgado H, Gama-Castro S, Martinez-Antonio A, *et al*. RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in Escherichia coli K-12. *Nucleic Acids Res* **2004**; 32: D303-306.
[341] Sivakumaran S, Hariharaputran S, Mishra J, Bhalla US. The Database of Quantitative Cellular Signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics* **2003**; 19: 408-415.
[342] Hucka M, Finney A, Bornstein BJ, *et al*. Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *Syst Biol (Stevenage)* **2004**; 1: 41-53.
[343] Hucka M, Finney A, Sauro HM, *et al*. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **2003**; 19: 524-531.
[344] Lloyd CM, Halstead MD, Nielsen PF. CellML. its future, present and past. *Prog Biophys Mol Biol* **2004**; 85: 433-450.
[345] Stromback L, Lambrix P. Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. *Bioinformatics* **2005**; 21: 4401-4407.
[346] Cao Y, Gillespie D, Petzold L. Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *J Comput Physics* **2005**; 206: 395-411.
[347] Cao Y, Gillespie DT, Petzold LR. Accelerated stochastic simulation of the stiff enzyme-substrate reaction. *J Chem Phys* **2005**; 123: 144917.
[348] Cao Y, Gillespie DT, Petzold LR. The slow-scale stochastic simulation algorithm. *J Chem Phys* **2005**; 122: 14116.
[349] Cao Y, Li H, Petzold L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J Chem Phys* **2004**; 121: 4059-4067.
[350] Machne R, Finney A, Muller S, Lu J, Widder S, Flamm C. The SBML ODE Solver Library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics* **2006**; 22: 1406-1407.
[351] Schmidt H, Jirstrand M. Systems Biology Toolbox for MATLAB. a computational platform for research in systems biology. *Bioinformatics* **2006**; 22: 514-515.
[352] Zi Z, Klipp E. SBML-PET. a Systems Biology Markup Language-based parameter estimation tool. *Bioinformatics* **2006**; 22: 2704-2705.
[353] Mulholland AJ. Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discov Today* **2005**; 10: 1393-1402.
[354] Urbanc B, Borreguero JM, Cruz L, Stanley HE. Ab initio discrete molecular dynamics approach to protein folding and aggregation. *Methods Enzymol* **2006**; 412: 314-338.
[355] Alvarez-Vasquez F, Sims KJ, Cowart LA, Okamoto Y, Voit EO, Hannun YA. Simulation and validation of modelled sphingolipid metabolism in Saccharomyces cerevisiae. *Nature* **2005**; 433: 425-430.
[356] Veflingstad SR, Almeida J, Voit EO. Priming nonlinear searches for pathway identification. *Theor Biol Med Model* **2004**; 1: 8.
[357] Kaufman M, Thomas R. Model analysis of the bases of multistationarity in the humoral immune response. *J Theor Biol* **1987**, 129: 141-162.
[358] Kaufman M, Thomas R. Emergence of complex behaviour from simple circuit structures. *C R Biol* **2003**; 326: 205-214.
[359] Thieffry D, Thomas R. Dynamical behaviour of biological regulatory networks--II. Immunity control in bacteriophage lambda. *Bull Math Biol* **1995**; 57: 277-297.
[360] Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, Palsson BO. Genome-scale metabolic model of Helicobacter pylori 26695. *J Bacteriol* **2002**; 184: 4582-4593.
[361] Schilling CH, Edwards JS, Letscher D, Palsson BO. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol Bioeng* **2000**; 71: 286-306.
[362] Schilling CH, Palsson BO. The underlying pathway structure of biochemical reaction networks. *Proc Natl Acad Sci USA* **1998**; 95: 4193-4198.
[363] Thieffry D, Thomas R. Qualitative analysis of gene networks. *Pac Symp Biocomput* **1998**: 77-88.
[364] Sorribas A, Hernandez-Bermejo B, Vilaprinyo E, Alves R. Cooperativity and saturation in biochemical networks: A saturable formalism using Taylor series approximations. *Biotechnol Bioeng* **2006.**
[365] Frasch HF, Barbero AM. Steady-state flux and lag time in the stratum corneum lipid pathway: results from finite element models. *J Pharm Sci* **2003**; 92: 2196-2207.
[366] Barrett CL, Kim TY, Kim HU, Palsson BO, Lee SY. Systems biology as a foundation for genome-scale synthetic biology. *Curr Opin Biotechnol* **2006**; 17: 488-492.
[367] Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* **2007**; 2: 727-738.
[368] Becker SA, Palsson BO. Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. *BMC Microbiol* **2005**; 5: 8.
[369] Becker SA, Price ND, Palsson BO. Metabolite coupling in genome-scale metabolic networks. *BMC Bioinformatics* **2006**; 7: 111.
[370] Duarte NC, Becker SA, Jamshidi N, *et al*. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA* **2007**; 104: 1777-1782.
[371] Duarte NC, Herrgard MJ, Palsson BO. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* **2004**; 14: 1298-1309.
[372] Duarte NC, Palsson BO, Fu P. Integrated analysis of metabolic phenotypes in Saccharomyces cerevisiae. *BMC Genomics* **2004**; 5: 63.
[373] Herrgard MJ, Covert MW, Palsson BO. Reconciling gene expression data with known genome-scale regulatory network structures. *Genome Res* **2003**; 13: 2423-2434.
[374] Herrgard MJ, Fong SS, Palsson BO. Identification of genome-scale metabolic network models using experimentally measured flux profiles. *PLoS Comput Biol* **2006**; 2: e72.
[375] Herrgard MJ, Lee BS, Portnoy V, Palsson BO. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome Res* **2006**; 16: 627-635.
[376] Joyce AR, Palsson BO. The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* **2006**; 7: 198-210.

[377] Joyce AR, Palsson BO. Toward whole cell modeling and simulation: comprehensive functional genomics through the constraint-based approach. *Prog Drug Res* **2007**; 64: 265, 267-309.

[378] Joyce AR, Reed JL, White A, Edwards R, *et al.* Experimental and computational assessment of conditionally essential genes in Escherichia coli. *J Bacteriol* **2006**; 188: 8259-8271.

[379] Price ND, Papin JA, Schilling CH, Palsson BO. Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol* **2003**; 21: 162-169.

[380] Reed JL, Palsson BO. Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome Res* **2004**; 14: 1797-1805.

[381] Reed JL, Patel TR, Chen KH, *et al*. Systems approach to refining genome annotation. *Proc Natl Acad Sci USA* **2006**; 103: 17480-17484.

[382] Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol* **2006**; 8: 1195-1203.

[383] Jahnke T, Huisinga W. Solving the chemical master equation for monomolecular reaction systems analytically. *J Mathematical Biol* **2007**; 54: 1-26.

[384] Munsky B, Khammash M. The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys* **2006**; 124: -.

[385] Qian H. Open-system nonequilibrium steady state: Statistical thermodynamics, fluctuations, and chemical oscillations. *J Phys Chem B* **2006**; 110: 15063-15074.

[386] Rao CV, Arkin AP. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *Journal of Chemical Physics* **2003**; 118: 4999-5010.

[387] Chance B, Greenstein DS, Higgins J, Yang CC. The mechanism of catalase action. II. Electric analog computer studies. *Arch Biochem* **1952**, 37: 322-339.

[388] Chance B, Greenstein DS, Roughton FJ. The mechanism of catalase action. I. Steady-state analysis. *Arch Biochem* **1952**, 37: 301-321.

[389] Michaelis L, Menten M. Die Kinetik der Invertinwirkung. *Biochem Z* **1913,** 49: 333-369.

[390] Savageau MA. Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation. *J Theor Biol* **1969**, 25: 370-379.

[391] Savageau MA. Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J Theor Biol* **1969,** 25: 365-369.

[392] Savageau MA. Biochemical systems analysis. 3. Dynamic solutions using a power-law approximation. *J Theor Biol* **1970**, 26: 215-226.

[393] Hlavacek WS, Savageau MA. Subunit structure of regulator proteins influences the design of gene circuitry: analysis of perfectly coupled and completely uncoupled circuits. *J Mol Biol* **1995**; 248: 739-755.

[394] Hlavacek WS, Savageau MA. Rules for coupled expression of regulator and effector genes in inducible circuits. *J Mol Biol* **1996**; 255: 121-139.

[395] Irvine DH, Savageau MA. Network regulation of the immune response: alternative control points for suppressor modulation of effector lymphocytes. *J Immunol* **1985**, 134: 2100-2116.

[396] Irvine DH, Savageau MA. Network regulation of the immune response: modulation of suppressor lymphocytes by alternative signals including contrasuppression. *J Immunol* **1985**, 134: 2117-2130.

[397] Savageau MA. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature* **1974,** 252: 546-549.

[398] Savageau MA. Optimal design of feedback control by inhibition. *J Mol Evol* **1974,** 4: 139-156.

[399] Savageau MA. Optimal design of feedback control by inhibition: dynamic considerations. *J Mol Evol* **1975**, 5: 199-222.

[400] Savageau MA. Mathematics of organizationally complex systems. *Biomed Biochim Acta* **1985**, 44: 839-844.

[401] Savageau MA. A theory of alternative designs for biochemical control systems. *Biomed Biochim Acta* **1985**, 44: 875-880.

[402] Savageau MA. The challenge of reconstruction. *New Biol* **1991**; 3: 101-102.

[403] Savageau MA. Reconstructionist molecular biology. *New Biol* **1991**; 3: 190-197.

[404] Savageau MA. Demand theory of gene regulation. I. Quantitative development of the theory. *Genetics* **1998**; 149: 1665-1676.

[405] Savageau MA. Demand theory of gene regulation. II. Quantitative application to the lactose and maltose operons of Escherichia coli. *Genetics* **1998**; 149: 1677-1691.

[406] Savageau MA. Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos* **2001**; 11: 142-159.

[407] Visser D, Heijnen JJ. The mathematics of metabolic control analysis revisited. *Metab Eng* **2002**; 4: 114-123.

[408] Visser D, Heijnen JJ. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metab Eng* **2003**; 5: 164-176.

[409] Heijnen JJ. Approximative kinetic formats used in metabolic network modeling. *Biotechnol Bioeng* **2005**; 91: 534-545.

[410] Alves R, Vilaprinyo E, Hernandez-Bermejo B, Sorribas A. Mathematical formalisms based on approximated kinetic representations for modeling genetic and metabolic pathways. *Biotechnol Genet Eng Rev* **2008,** 25: submitted.

[411] Liebermeister W, Klipp E. Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor Biol Med Model* **2006**; 3: 41.

Received: February 13, 2007 Revised: February 20, 2008 Accepted: March 10, 2008