Advancing drug discovery through systems biology

Eugene J. Davidov, Joanne M. Holland, Edward W. Marple and Stephen Naylor

Pharmaceutical companies are facing an urgent need to both increase their lead compound and clinical candidate portfolios and satisfy market demands for continued innovation and revenue growth. Here, we outline an emerging approach that attempts to facilitate and alleviate many of the current drug discovery issues and problems. This is, in part, achieved through the systematic integration of technologies, which results in a superior output of data and information, thereby enhancing our understanding of biological function, chemico–biological interactions and, ultimately, drug discovery. Systems biology is one new discipline that is positioned to significantly impact this process.

Eugene Davidov*
Joanne M. Holland
Edward W. Marple
Stephen Naylor
Beyond Genomics
40 Bear Hill Road
Waltham, MA 02451, USA
tel: +1 781 434 0225
fax: +1 781 895 1119
*e-mail: edavidov@beyondgenomics.com

Systems biology uses an integrated approach to study and understand the function of biological systems, and how perturbations of such systems, for example the administration of a therapeutic drug, affect their function. The biological system can be at the level of a subcellular organelle, cell, organ, tissue or organism. The approach requires the simultaneous static and/or temporal measurement of genomic, proteomic and metabolomic parameters. Furthermore, it can only be successfully applied with a seamlessly integrated bioanalytical and computational biology capability in place. Here, we outline general approaches to the study of biological complexity through systems biology, and provide examples of successful application of this discipline in the understanding of disease.

Until recently, drug discovery has primarily been a linear process based on the sequential approaches of biology and chemistry. This has led to the separation of scientific disciplines into 'functional silos', with relatively limited cross-talk within the discovery process. As a consequence, the primary approach in the drug discovery process typically involves screening vast, randomized chemical libraries against a small number of pharmacologically relevant, and in some instances poorly defined, biological targets [1]. Although this approach has provided some success, the impact of HTS, ultra-HTS and high-speed combinatorial chemistry technologies has been less than the initially projected 'several-fold' increase in drug discovery productivity [2]. This is best reflected in the relationship between the estimated number of HTS assays per target versus the number of new chemical entities (NCEs) reaching the market. In the past decade, this 'numbers game' indicated a trend of diminishing returns where assays per target have increased exponentially from several thousand to hundreds of millions. At the same time, the number of resultant NCEs remained stagnant [3,4]. Nevertheless, several specific molecular targets have been identified and exploited for the treatment of a broad range of pathogenic conditions, including: β-adrenoceptor antagonists and angiotensin-converting enzyme inhibitors for cardiac arrhythmias; HMG-CoA reductase inhibitors (statins) for hyperlipidemia; and cyclooxygenase-2 (COX-2) inhibitors for arthritis and general inflammation. However, in multifactorial diseases, where multiple targets or pathways have to be affected for successful treatment outcomes, linking structurally and functionally characterized targets with the disease still remains a challenge.

A new knowledge-based approach has emerged that is a more comprehensive, systems biology-based approach to biological function, cellular processes and disease-mediated processes, and that increases the probability of success in the drug discovery process. The emphasis is on the integration of analytical technologies and information, and includes the incorporation of structural

1359-6446/03/$ – see front matter PII: S1359-6446(03)02600-X

data with specific biological pathway information; for example, the synthesis of structurally defined chemical libraries that target selected protein families such as kinases, phosphatases and G-protein-coupled receptors (GPCRs). Systems biology signals a departure from the now common view in drug discovery of 'single target, one drug, lone therapeutic indication'. Targeting a broader range of related biological structures should result in compounds that have common structural and functional properties, and common mechanisms of action, ultimately creating the potential for the application of a therapeutic to multiple diseases by targeting common pathways implicated in pathogenesis.

The impetus for paradigm shift

As the discovery process shifts to a focused high-throughput biological mode, that is, global gene expression analysis and whole-genome functional analysis [5–9], the sequential nature of the established drug discovery process will become less viable. This reflects an understanding that many genetic and metabolic disorders, such as cancer, Alzheimer's disease and atherosclerosis, are caused and mediated by complex multi-molecular interactions that cannot be readily explained by an alteration in a single gene, gene product, or enzymatic cascade. Therefore, although we have a greater understanding of molecular and cellular processes, fewer than 500 validated targets have been exploited for therapeutic intervention to-date [10].

Genes that interact to produce multifactorial disease phenotypes present many new attractive targets, effectively increasing by several-fold the biological structural space that needs to be explored for drug discovery. For example, novel compounds that target the GPCR family could potentially have dramatic therapeutic benefit in a broad spectrum of diseases. However, prospects for multiple validated targets for specific conditions in this field are rare. This limitation is somewhat compelling, as illustrated by the competition among pharmaceutical companies for the same therapeutic indications using similar drugs against the same target family; for example, serotonin-5HT_2 for CNS disorders, and histamine H_2 for peptic ulcers [11]. The GPCRs are particularly attractive for pathway-based drug discovery as they are key facilitators of cellular signaling cascades. Unraveling the complex physiology of receptor-mediated signaling, and linking these signaling networks to disease presents exciting opportunities in the identification of multiple new targets, and creates new and viable options for therapy.

As noted previously, systems biology is an emerging and promising discipline that aims to facilitate the understanding of phenotypic variation and build comprehensive models of cellular organization and function. Such an approach enables the integration of massive quantities of complex data generated by genomic, proteomic and metabolic analyses, and provides an interactive process to translate the findings into novel therapies. Herein lies the opportunity for companies with cutting-edge high-throughput 'omics' platforms and integrated informatics capabilities to create medicines of the future and to secure solid foundations in a new paradigm of drug discovery. This paradigm is underscored by the increasing number of commercial and academic organizations that have staked their future on the growing recognition and use of system-based discovery platforms (Table 1). Recent announcements confirm that big pharma is joining ranks with advocates of the systems biology approach. Eli Lilly (http://www.lilly.com/) has announced the establishment of a new Center for Systems Biology in Singapore accompanied by an investment of US$140 million.

Systems biology: the emerging discipline

Researchers looking at information on a single level, for example, a DNA expression profile, only observe a partial composite of the biological system. Therefore, accumulating information at multiple levels (e.g. genes, proteins and metabolites), and studying complex relationships among such molecules, reveals information about pathways and ultimately helps to focus in on and better define therapeutic targets [12]. This requires a novel unified, holistic system-level approach to define the relationship between the genotype, phenotype and drug. Systems biology facilitates the understanding of how very complex and dynamic systems work, thereby providing insight into both the underlying causes of pathogenic changes and the options available to treat the whole disease rather than one specific symptom.

The idea of system-level analysis has been around since Norbert Weiner introduced mathematical models of the control and communication of complex systems [13], and Ludwig von Bertalanffy introduced the General Systems Theory [14] over three decades ago. By emphasizing the importance of 'wholeness', the concepts and central themes postulated by contemporary systems biology approaches are proposed to transform the discovery process by the parallel study of complex relationships among genetic, proteomic and metabolic networks.

Genomic technologies have generated vast amounts of biological data that need to be assembled and defined in a way that accurately describes living organisms. Many of the gene, protein, small-molecule and interaction databases, including the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov), Protein Data

Table 1. Organizations that incorporate systems biology approaches for discovery efforts (representative selection)

Name	Approach
Bioseek (http://www.bioseekinc.com)	Uses systems biology approach to study primary human cell disease models
Beyond Genomics (http://www.beyondgenomics.com)	Technology platform facilitates analysis of clinically relevant samples and integrates data from the gene, protein, metabolite and clinic for biomarker and target identification
Cellnomica (http://www.cellnomica.com)	Conducts novel multicellular modeling in drug discovery and development
Cellzome (http://www.cellzome.com)	Proprietary functional proteomics technology for therapeutic target discovery, validation and drug development
Department of Energy's Genomes to Life initiative (http://doegenomestolife.org/overview.pdf)	The Genomes to Life roadmap (plans to design and exploit new high-throughput strategies to obtain a blueprint of how living systems function)
Eli Lilly Center for Systems Biology (http://www.lilly.com)	Focuses on integration of proteomic and genomic technologies to support drug discovery efforts
Entelos (http://www.entelos.com)	Biosimulation company that develops computer models of human disease using novel PhysioLab® technology
Institute for Systems Biology (http://www.systemsbiology.org)	Broad based program. Uses systems biology to investigate the complex interaction of biological elements that form hierarchical networks that define systems
Kitano Symbiotic Systems Project (http://www.symbio.jst.go.jp)	The project aims to understand and design biological systems, thus creating a new paradigm in biology Focuses on model organisms including fruit fly, yeast and bacteria.
Physiome Sciences (http://www.physiome.com)	Biosimulation company that has created and develops integrated software platform for computer-based biological models applicable to drug discovery
SurroMed (http://www.surromed.com)	Develops and implements biological marker discovery platform to profile biochemical components in blood and other biological samples

Bank (http://www.rcsb.org/pdb), Kyoto Encyclopedia of Genes and Genomes (http://www.genome.ad.jp/kegg), and the Biomolecular Interaction Network Database (http://www.isc.org/products/BIND) (see Table 2), depict modular or static biochemical states. Progressive analytical and mathematical tools are needed to integrate disjointed biological events. For example, activation of selected cell-surface receptors in response to external stimuli, or the breakdown in the apoptosis cascade associated with tumorogenesis are dynamic events, and knowledge-based models are needed to explain such pathological processes underlying pleiotropic disease. Currently, several concepts of systems biology exist, each attempting to uniquely equip its practitioners with a specific road-map for system-level understanding. Some systems biology approaches, such as *in silico* modeling, represent highly approximated and constrained views of system infrastructure and function, principally using known relationships to create virtual systems. By contrast, approaches proposed by the Kitano systems biology project [Exploratory Research for Advanced Technology (http://www.cds.caltech.edu/erato)], Japan Science and Technology Corporation (http://www.jst.go.jp/EN) and Hood (Systems Biology Institute, http://www.systemsbiology.org) advocate the use of systems biology general operators, such as system structure, dynamics, control methods and design methods, to specific areas of biology [15,16]. According to Kitano, the pivotal process in systems-level analysis is the comprehensive and precise measurement of all components in a system and their functional outputs within the cellular space. The ultimate outcome of such large-scale measurements is the creation of high-resolution cellular simulation models, supported and refined by iterative processes of hypothesis-driven- and wet-experimentation.

An intriguing and somewhat different concept looks for commonalities in design between biological systems and the complex organization circuitry found in technology. Arguably, biology and technology differ on many levels. The similarities, however, emerge through concepts of convergent evolution, modularity, and elementary feedback-control, which are well-established processes in biology and engineering (see Box 1). This approach draws comparisons between nature- and human-conceived designs to reveal a blueprint of a system. Allometric characteristics (see Box 1), for example, of a fruit fly versus an aircraft, or complex circuit networks in electronic devices versus cellular

Table 2. Structural and/or functionally curated databases and biomolecular interactions resources

Name	Category	Web Site
BIND (Biomolecular Interaction Network Database)	Molecular interaction network database	http://www.binddb.org/
DIP (Database for Interacting Proteins)	Protein-protein interactions	http://dip.doe-mbi.ucla.edu/
EMP (Enzymes and Metabolic Pathways)	Enzymes and metabolic pathways maps	http://emp.mcs.anl.gov/
GO (Gene Ontology)	Dynamic controlled vocabulary for knowledge assembly	http://www.geneontology.org/
KEGG (Kyoto Encyclopedia of Genes and Genomes)	Metabolic pathways	http://www.genome.ad.jp/kegg/
LIGAND (Chemical database for enzyme reactions)	Chemical compounds and reactions in biological pathways	http://www.genome.ad.jp/ligand/
NCBI (National Center for Biotechnology Information)	Complete genomes and analysis	http://www.ncbi.nlm.nih.gov
SWISS-PROT	Protein sequence database	http://www.ebi.ac.uk/swissprot/
TIGR (The Institute for Genomic Research)	Comprehensive microbial resource	http://www.tigr.org
TRANSFAC (Transcription factor database)	Transcription regulation	http://transfac.gbf.de/TRANSFAC/
WIT (What Is There)	Metabolic reconstruction	http://wit.mcs.anl.gov/WIT2/

Box 1. Terms and definitions

Allometric
The regular systematic pattern of growth such that the mass or size of any organ or part of a body changes in shape in response to size changes. This can be expressed in relation to the total mass or size of the entire organism according to the allometric equation:

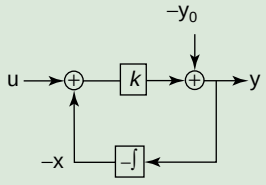
$$Y = bx^{\alpha} \qquad \text{[Eqn 1]}$$

where Y = mass of the organ, x = mass of the organism, α = growth coefficient of the organ, and b = a constant. Allometric scaling is common in nature, both when comparing two animals of two different sizes and when comparing the same animal at two different sizes (i.e. growth).

Convergent evolution
The development of superficially similar structures in unrelated organisms or biological pathways, usually because the organisms or pathways evolved in the same kind of environment. Examples include the wings of insects and birds, and the streamlined bodies of whales and fish.

Modularity
A design concept in engineering that enables engineers to build complex systems out of simpler modules, which can be tested independently before being integrated. As an additional benefit, such modules can be used in more than one system, saving development time and increasing quality. Modules also make maintenance of such systems much easier. These two aspects increase security with such complex systems through ease-of-testing during design and by facilitating failure detection and correction during operation. Modularity also appears to have applications in biology. There are metabolic units that can be found in different organisms, and homologies are used to compare the genomes of different organisms to gain insights into development, physiology and evolution.

Elementary feedback control
A primary role of feedback control is to ensure robust signal tracking and to decrease uncertainty owing to noise and other disturbances in a system. A schematic example of a feedback control loop is shown below:



where $x = y$; $y = y_1 - y_o = k(u - x) - y_o$; $y(t) \rightarrow 0$ as $t \rightarrow \infty$ if $k>0$

A block diagram of integral feedback control. The variable u is the input for a process with gain k. The difference between the actual output y and the steady-state output y_0 represents the normalized output or error, y. Integral control arises through the feedback loop in which the time integral of y, x is fed back into the system. This results in $\alpha = y$ and $y = 0$ at steady-state for all u [a].

References

a Yi, T-M. *et al.* (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4649–4653
b A Glossary for Systems Biology, available at: http://www.sysbio.de/projects/glossary

signal transduction cascades, provide common blueprints of system modules and whole-system regulatory architectures, which can be reverse-engineered [17–19]. Essentially, biological networks, such as gene or protein regulatory circuits, operate as patterns of interconnections that define the overall network. Specific patterns or 'motif' occurrences in organized complex networks evolve at a much higher frequency than those in randomized systems [20]. One can infer a set of rules that govern connection patterns within and across network modules by comparison of organized and randomized systems. Subsequent assembly of these modules into an operational system is at the center of reverse engineering and is complementary to the experimental, statistical and mathematical methods that comprise the systems biology tool chest.

Systems biology: metabolic connections

Metabolic engineering provides an example of altering a system to achieve a desired physiological change. The methods used in metabolic engineering are relatively mature and have been used extensively in the fields of industrial microbiology and microbial physiology. Classical strategies involve strain-improvement screens for enhancing fermentation properties, and selection for high producers of natural products (i.e. penicillin) [21]. Genomics has created new opportunities in applied metabolic design, including targeted approaches through direct DNA manipulation to attain preferred phenotypes. An example of such rational engineering includes industrially and therapeutically important recombinant protein production. Functional genomics and proteomics approaches in conjunction with metabolic control analysis (MCA) [22] are increasingly used to study the metabolic status of living cells (Box 2). As an emerging science, metabolomics relies on the systematic comparative analysis of biological samples both to determine the physiological status of the sample, and to understand how specific perturbations at the gene and protein levels can be related to changes in metabolic flux and enzymatic product:substrate ratios [23,24].

Defining a system by the dynamic state of its metabolome relies on the effective integration of omics data because the metabolic state of the system is largely derived from the global expression of its genome and proteome [25]. As with gene expression and proteomic analysis, interpretation of complex metabolic patterns is usually aided by an array of multivariate statistical tools. This strategy has been used to reveal functional differences of silent phenotypes in the sugar phosphate conversion pathway in *Saccharomyces cerevisiae* [26]. These methodologies depend principally on the accurate quantitative pattern analysis of metabolic constituents in the cell or cellular compartments.

Box 2. Metabolomics

Metabolic control analysis (MCA) is a method for analyzing how the control of fluxes and intermediate concentrations in a metabolic pathway is distributed among the different enzymes that constitute the pathway. Instead of assuming the existence of a unique rate-limiting step, it assumes that there is a definite amount of flux control and that this is spread quantitatively among the component enzymes.

Medium- and high-resolution methods, such as NMR, GC–MS and LC–MS, are often the bioanalytical techniques of choice in metabolite quantitative analysis [27,28]. The primary result of such measurements is a metabolic profile that serves as a reporter for the biochemical status of the system and can be instrumental in identifying key enzymatic steps in metabolically controlled pathways. Physicochemical parameters, such as rate-limiting steps, stoichiometric relationships, kinetic and control coefficients, can be obtained from metabolic-flux measurements and used in the modeling of metabolic networks.

Several approaches have been reported outlining computational strategies to predict reaction sequence and pathway modulations in *E. coli* in response to genetic perturbations and environmental challenge [29,30]. In the case of metabolic regulation, the information gained from such procedures augments and refines knowledge provided by gene and protein analysis. Obtaining detailed metabolic maps of tissues, such as liver or kidney, can have a dramatic impact on the drug development process. The value of metabolic system analysis has been made clear through the discovery of diagnostic and prognostic biomarkers [31], mechanisms of drug action [32], efficient management of absorption, distribution, metabolism, excretion (ADME) [33], and prediction of drug toxicity [34]. Metabolic approaches in systems biology typically combine high-throughput profiling of endogenous metabolites, MCA, statistical analysis and *in silico* methods to define phenotypic variation by reversing the established genetic analysis sequence; that is, establishing links between genotype and metabolome.

In silico world

In silico biology and its associated approaches have increasingly been used in the analysis of biological processes [35,36]. Most *in silico* models incorporate real biological data derived from the biochemical properties of gene products. The data are converted to a numerical format and plugged into a set of equations and algorithms that

Table 3. Organizations developing *in silico* tools and descriptors for general and specific system modeling

Project/Organization	Brief description
Alliance for cellular signaling (http://www.cellularsignaling.org)	Identify all the proteins that comprise the signaling systems. Assess the time-dependent information flow through the systems in both normal and pathological states
CellML™ Physiome Sciences (http://www.physiome.com/) Bioengineering institute at university of Auckland (http://www.bioeng.auckland.ac.nz/home/home.php)	CellML™ is an XML-based mark-up language designed to facilitate exchange and integration of biological models
E-Cell Project Keio University (http://www.e-cell.org/about/index.htm)	A modeling and simulation capability for biological processes
SBML – Systems Biology markup language Caltech ERATO Kitano Systems Biology project (http://xml.coverpages.org/sbml.html)	Develop the Systems Biology Markup Language (SBML) to represent and model information components in the system
Virtual Cell University of Connecticut Health Center (http://www.nrcam.uchc.edu)	Virtual Cell environment can be applied to mammalian cells based on precise measurement of how molecules diffuse to react with target cells

attempt to describe the system to be simulated. The model can be further refined by perturbing the model system in ways that approximate genetic alteration or the effects of drug action. Advances in mathematical modeling tools and computer simulation have positioned *in silico* biology to capitalize on technological breakthroughs across all sectors of biomedical research, from target selection to the emerging field of personalized medicine. As the starting data for simulation and modeling comprise heterogeneous data types (gene expression, protein function and metabolic flux), the inter-relationships among molecular components in a target tissue or organ are highly dynamic and nonlinear. Given this complex relationship, cellular processes are not easily amenable to a specific single-modeling approach or solution. The basis of every successful model is the source and the quality of data. The data format, however, is equally important in the context of seamless information exchange between various simulation and modeling software environments. Table 2 lists several databases that contain annotations of functional and interaction data applicable to computational analysis.

Two principal modeling approaches have emerged, and are referred to as mechanistic or data-driven, and qualitative or hypothesis-driven modeling [36,37]. Mechanistic models often rely on experimental data input, typically from high-throughput omics technologies. Modeling various aspects of biological processes using mechanistic approaches can potentially enable the linking of specific regulatory nodes and pathways within cells and tissues to underlying causes of disease, thereby providing highly specific targets for novel therapies. Owing to signal detection limitations of the current analytical technologies, and the sheer number and complexity of possible interactions within a system, mechanistic models often fail to provide complete solutions for a broad range of cellular processes. Hypothesis-driven simulation bridges gaps in available data to construct logical models of selected biological phenomena that fit known information. Presently, several organizations are developing novel tools to formulate common descriptors for biological systems modeling (Table 3).

Specific qualitative models have been developed by the incorporation of Boolean and fuzzy-logic rule-based approaches [38]. For example, gene networks can be explored using Boolean parameters that assign simple 'on–off' states to individual connections. However, the majority of biological processes and pathways cannot be readily described by basic 'true–false' values. Fuzzy logic (Box 3) represents an extension of traditional Boolean approaches. It is being increasingly used in computational biology to model gene, protein and metabolic networks in a more realistic fashion. If sufficient amounts of data are available – in combination with temporal information – the increasing levels of complexity (from subcellular mechanisms to whole cell, organs and, ultimately, whole body) can be simulated through statistical influence models that might include neural and Bayesian networks. When combined with the development of more advanced computational tools, these new mathematical and statistical approaches might, in the future, become central enabling technologies in drug discovery.

Applications in drug discovery

The impact of systems biology on drug discovery can be realized in several fundamental ways. Initially, systems

biology can be used to identify new uses for existing molecular targets. This is exemplified by characterized proteins that have previously lacked an established connection to a specific disease. A second use is the identification of novel molecular targets with a connection to disease. In this case, novel molecular targets might be endogenous proteins not previously characterized and therefore do not have an established connection to disease. Alternatively, they might be mutant proteins (inherent to the disease) not previously identified. Here, both objectives might be achieved through a systems biology analysis of the differences between normal and diseased samples, as discussed later. A third approach in which systems biology might be applied to drug discovery is in deciphering complex signaling relationships, which would enable targeting of the most appropriate region of a signaling cascade for the development of more efficient and safe therapeutic agents. In addition to the differential analysis of diseased and normal samples mentioned previously, this final use might also be accomplished through studies that characterize perturbations of the biological system caused by small molecules, including existing therapeutic agents. This will enable researchers to discriminate between cellular changes associated with therapeutic benefits and cellular changes associated with side-effects: information relevant for development and regulatory purposes, and for marketing the therapeutic.

To illustrate the application of systems biology to the first two potential uses described earlier, consider the information that could be derived from the proteomic and metabolomic analysis of a hypothetical diseased state that is ultimately attributed to altered expression of a previously uncharacterized kinase. A proteomic analysis alone might identify several proteins with distorted expression levels in the diseased sample with respect to the normal sample. Initially, each of these proteins represent potential causative elements of the disease. Alternatively, metabolomic analysis alone might uncover prominent alterations in protein phosphorylation levels in the diseased sample, suggesting the involvement of altered kinase or phosphatase activity. However, an integrative systems biology approach, in which both proteomic and metabolomic data acquisition and analysis are performed in parallel, enables a rapid link of the particular kinase to the disease in question. The power inherent in the combination and integration of multiple parallel approaches is the essence of systems biology.

Decades of genomic, biochemical and cell biology research have provided a wealth of information regarding the molecular basis of cell function. In many diseases, proteins linked to the disorder – which ultimately become molecular targets for drug discovery – might be proteins that are already well-characterized but whose connection to the disease has not previously been established. Systems biology analysis can facilitate the elucidation of such connections, providing the opportunity for new uses of existing targets. A successful example of exploiting a known target with a novel connection to a therapeutic endpoint is the discovery and successful development of sildenafil (Viagra™; Pfizer, http://www.pfizer.com) [39]. Pfizer's program for the discovery of selective phosphodiesterase (PDE) type 5 inhibitors was initiated with the intention of attaining therapeutic agents for use in cardiovascular disorders. The ultimate connection of this molecular target to mechanisms of smooth muscle relaxation provided the basis of investigations of these agents in the treatment of male erectile dysfunction, ultimately yielding sildenafil.

Despite our wealth of biological knowledge, it has been recognized that many causative agents for disease might be novel, including wild-type or mutant proteins not previously identified or characterized. Systems biology can serve to elucidate these causative agents of disease, thereby identifying completely new targets for drug discovery. The delivery of such novel targets to discovery programs often provides the first opportunity for the development of adequate therapies. The importance of

Box 3. Logic in computational biology

Fuzzy logic is a departure from classical two-valued sets and uses 'soft' linguistic system variables (e.g. large, hot, tall) and a continuous range of truth values in the interval [0,1], rather than strict binary (true or false) decisions and assignments and basic Boolean operators (AND, OR and NOT). Formally, fuzzy logic is a structured, model-free estimator that approximates a function through linguistic input–output associations. Fuzzy rule-based systems apply these methods to solve many types of real-world problems, especially where a system is difficult to model.

Boolean algebra

YES or NO logic (0 or 1)
Something is either 'part of A' or 'not A'
It cannot be 'A' and 'not A' at the same time
Does not effectively mimic human thinking

Fuzzy logic

Fuzzy sets
Something can be part of A and 'part of not A' at the same time
Mimics human thinking and decisions

connecting a molecular target to disease, the ability of drug discovery programs to exploit the connections for the development of therapies, and the subsequent benefit in terms of human health outcome, was most recently exemplified by the success story of Gleevec™ (Novartis, http://www.novartis.com) [40]. The discovery of the bcr–abl gene, which encodes a protein with elevated tyrosine-kinase activity, provided a drug target with differential activity between normal and leukemic cells. Novartis' drug discovery program focused on the development of inhibitors versus this target, and ultimately produced Gleevec™, the first selective tyrosine-kinase inhibitor to be approved for the treatment of cancer.

The third use of systems biology mentioned earlier, that is, the ability to decipher complex inter- and intra-cellular signaling relationships, should significantly enhance discovery efforts. By defining the behavior of entire signaling networks, the researcher has the capacity to focus on the most appropriate region of a cascade for the development of efficacious and safe therapeutic agents. Signaling events within or between cells are not restricted to linear pathways, but are well-known to be complex and dynamic networks [41]. Systems biology is uniquely positioned to work within these complexities as its very essence is the comprehensive analysis of an entire system. Given some knowledge of signaling events associated with treatment of a particular disease, it might be possible to exploit a systems biology-based effort to revise the entry point within the system to optimize efficacy and/or minimize side-effects. The necessity for this kind of approach is perhaps best exemplified by efforts in the field of depression [42]. Although the majority of research within this field has focused on the monoamine hypothesis, the complete pharmacological basis for this disorder remains poorly understood. This is exemplified by shifts in discovery programs from those targeting selective serotonin reuptake inhibitors (SSRIs) to those targeting a variety of postulated molecular connections, including NMDA, neuropeptide, nicotinic and cannabinoid receptors. Despite many entries of novel therapeutics onto the market, the elimination of side-effects leading to non-compliance, and the problems associated with delayed onset to efficacy, have yet to be solved.

One way in which we are able to acquire an understanding of signaling networks is through the comparative analysis of diseased and normal samples. An alternative approach is through drug perturbation studies. Specifically, a comprehensive analysis of a biological system upon treatment with small molecules might define the molecular consequences affected by these molecules. In many instances, existing therapies operate via molecular targets whose connection to the desired therapeutic outcome is coupled with connections to unwanted side-effects. Appropriate selection of small molecules for parallel perturbation studies enables the researcher to identify differences invoked by agents known to elicit side-effects and those exclusively providing efficacy. Identification of points of entry to the cascade where molecular targets are free of the connection to side-effects could provide dramatic improvements in existing therapies.

Central to the successful implementation of systems-level research is the rigorous selection process and preparation of disease-relevant clinical samples. The comparative nature of the integrated analysis dictates that samples fall into well-defined phenotypic categories, characterized either by disease pathology or synthetic perturbation (e.g. genetic mutations and exposure to drugs). Clearly, exhaustive bioanalytical measurements must be taken across all static and functional components of the system. Recent studies have shown interesting results from partial parallel analyses of genetically perturbed organisms by correlating gene expression with protein function or physiological profiles [43,44]. However, to take full advantage of multi-tiered quantitative analysis, a platform needs accounting tools for the metabolic components and sophisticated statistical tools to integrate data. One difficulty in integrating heterogeneous datasets is adapting a compatible format for the entire signal-based measured output of the system (e.g. spectral intensity, fluorescence intensity, concentration, and digital image). Systems biology approaches that interface data-generating technology with customized statistical analysis and normalization tools enable high-resolution modeling and knowledge assembly of integrated biochemical data [45]. A recent study demonstrating the use of this platform describes targeted pattern recognition analysis for the elucidation of the causes of lipoprotein metabolism dysregulation in a well-established model of atherosclerosis (E. Davidov *et al.*, unpublished results) Comprehensive bioanalytical strategies for simultaneous measurements of transcript, protein and metabolite content were successfully implemented in the study. To extract maximum value from chemical component correlations established by statistical inference, the study used a combination of informatics tools, collectively termed BioSystematics™ (Beyond Genomics, http://www.BeyondGenomics.com). The non-linear statistical and ontology-based data-mining algorithms of this platform enable accurate predictions of the biological significance of measured correlations in proposed pathway topologies. It shows the direct, simultaneous correlation of genes, proteins, and metabolites, with the ultimate goal of ascertaining causal relationships.

The ability of systems biology to impact the drug discovery and development process from target identification through to clinical development is illustrated in Fig. 1. The intrinsic value of systems biology is, in essence, its ability to unite individual fields of research devoted to structural, functional and dynamic aspects of biology into one powerful discipline.

Concluding remarks

The integrative biology approaches discussed here hold great potential for the future of drug discovery and the overall understanding of biological phenomena. The combination of mature and novel genomic technologies, improved data management, and mining environments in conjunction with *in silico* methods and clinically relevant samples, positions systems-based analysis at the frontline of medical research. The deliverables for systems biology will span the spectrum of the discovery process, from drug development and clinical trials to personalized medicine. As it matures, this discipline promises to become a dominant approach in drug discovery and development by overcoming the limitations of individual omics technologies.

References

1 Wess, G. (2002) How to escape the bottleneck of medicinal chemistry. *Drug Discov. Today* 7, 533–535
2 Manly, C.J. (2001) Discovery technologies and the issues facing drug discovery. *Curr. Drug Discov.* (September), 9–10
3 Gaudillière, B. *et al.* (2000) To market, to market – 2000. *Annu. Rep. Med. Chem.* 36, 293–318
4 Oprea, T.I. (2002) Chemical space navigation in lead discovery. *Curr. Opin. Chem. Biol.* 6, 384–389
5 Schena, M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470
6 Templin, M.F. *et al.* (2002) Protein microarray technology. *Trends Biotechnol.* 20, 160–166
7 Kuruvilla, F.G. *et al.* (2002) Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. *Nature* 416, 653–657
8 Hieter, P. and Boguski, M. (1997) Functional genomics: it's all how you read it. *Science* 278, 601–602
9 Zheng, X.F.S. and Chan, T-F. (2002) Chemical genomics and the global study of protein functions. *Drug Discov. Today* 7, 197–205
10 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–964
11 Wise, A. *et al.* (2002) Target validation of G-protein coupled receptors. *Drug Discov. Today* 7, 235–246
12 Bonetta, L. (2002) Systems biology – the new R&D buzzword? *Nat. Med.* 8, 315–316
13 Weiner, N. (1948) *Cybernetics or Control and Communication in the Animal and Machine*, MIT Press
14 Bertalanffy, L.V. (1969) *General System Theory, Foundations, Development, Applications*, George Brazziller, New York
15 Kitano, H. (2001) *Foundations of Systems Biology*. (Kitano, H. ed.), MIT Press
16 Kitano, H. (2002) Systems biology: a brief overview. *Science* 295, 1662–1664
17 Csete, M.E. and Doyle, J.C. (2002) Reverse engineering of biological complexity. *Science* 295, 1664–1669
18 Whitfield, J. (2001) All creatures great and small. *Nature* 413, 342–344
19 Freeman, M. (2000) Feedback control of intercellular signalling in development. *Nature* 408, 313–319
20 Milo, R. *et al.* (2002) Network motifs: simple building blocks of complex networks. *Science* 298, 824–827
21 Olsson, L. (2002) An expanded role for microbial physiology in metabolic engineering and functional genomics: moving towards systems biology. *FEMS Yeast Res.* 2, 175–181
22 Cascante, M. *et al.* (2002) Metabolic control analysis in drug discovery and disease. *Nat. Biotechnol.* 20, 243–249
23 Sanford, K. *et al.* (2002) Genomics to fluxomics and physiomics – pathway engineering. *Curr. Opin. Microbiol.* 5, 318–322
24 Fiehn, O. *et al.* (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18, 1157–1161
25 Glassbrook, N. *et al.* (2000) Metabolic profiling on the right path. *Nat. Biotechnol.* 18, 1142–1143
26 Raamsdonk, L.M. *et al.* (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat. Biotechnol.* 19, 45–50
27 Nicholson, J.K. *et al.* (1999) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 29, 1181–1189
28 Roessner, U. *et al.* (2000) Technical advance: simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J.* 23, 131–142
29 Verner, J. (2000) Large-scale prediction of phenotype: concept. *Biotech. Bioeng.* 69, 664–678
30 Burgard, A.P. *et al.* (2001) Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnol. Prog.* 17, 791–797
31 Griffin, J.L. *et al.* (2002) Metabolic profiles of dystrophin and utrophin expression in mouse models of Duchenne muscular dystrophy. *FEBS Lett.* 530, 109
32 Sikora, K. (2002) Surrogate endpoints in cancer drug development. *Drug Discov. Today* 7, 951–956
33 Butina, D. *et al.* (2002) Predicting ADME properties *in silico*: methods and models. *Drug Discov. Today* 7 (Suppl.), S83–S88
34 Reo, N.V. (2002) NMR-based metabolomics. *Drug Chem. Toxicol.* 25, 375–382
35 Palsson, B. (2000) The challenges of *in silico* biology. *Nat. Biotechnol.* 18, 1147–1150
36 Arkin, A.P. (2001) Synthetic cell biology. *Curr. Opin. Biotechnol.* 12, 638–644
37 Somogyi, R. and Greller, L.D. (2001) The dynamics of molecular networks: applications to therapeutic discovery. *Drug Discov. Today* 6, 1267–1277
38 Thieffry, T.D. (1998) Qualitative analysis of gene networks. In *Pacific Symposium on Biocomputing*, pp. 77–88, World Scientific Publishers, Singapore
39 Terret, N.K. *et al.* (1996) Sildenafil (Viagra™), a potent and selective inhibitor of type 5 cGMP phosphodiesterase with utility for the treatment of male erectile dysfunction. *Bioorg. Med. Chem. Lett.* 6, 1819
40 Capdeville, R. *et al.* (2002) Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discov.* 1, 493–502
41 Marinissen, M.J. and Gutkind, J.S. (2001) G-protein-coupled receptors and signaling networks: emerging paradigms. *Trends Pharmacol. Sci.* 22, 368–376
42 Macritchie, K.A.N. and Young, A.H. (2001) Emerging targets for the treatment of depressive disorder. *Expert Opin. Ther. Targets* 5, 601
43 Ideker, T. *et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934
44 Stoll, M. *et al.* (2001) A genomic-systems biology map for cardiovascular function. *Science* 294, 1723–1726
45 Neumann, E. and Thomas, J. (2002) Knowledge assembly for the life sciences. *Drug Discov. Today* 7 (Suppl.), S160–S162