nature biotechnology

PERSPECTIVE

Systems biology in drug discovery

Eugene C Butcher[1,2], Ellen L Berg[3] & Eric J Kunkel[3]

The hope of the rapid translation of 'genes to drugs' has foundered on the reality that disease biology is complex, and that drug development must be driven by insights into biological responses. Systems biology aims to describe and to understand the operation of complex biological systems and ultimately to develop predictive models of human disease. Although meaningful molecular level models of human cell and tissue function are a distant goal, systems biology efforts are already influencing drug discovery. Large-scale gene, protein and metabolite measurements ('omics') dramatically accelerate hypothesis generation and testing in disease models. Computer simulations integrating knowledge of organ and system-level responses help prioritize targets and design clinical trials. Automation of complex primary human cell–based assay systems designed to capture emergent properties can now integrate a broad range of disease-relevant human biology into the drug discovery process, informing target and compound validation, lead optimization, and clinical indication selection. These systems biology approaches promise to improve decision making in pharmaceutical development.

Drug discovery and systems biology began together: in traditional or 'folk' medicine, herbal drugs were discovered through direct if anecdotal observations in people with diseases, the most relevant complex biological systems there are. With the advent of chemistry in the late 1800s and early 1900s, derivatives of natural products and subsequently novel synthetic chemicals made their way into drug discovery pipelines; but screening was still in the setting of complex disease biology, with animals replacing patients as the primary 'guinea pigs.' Most of today's pharmaceuticals (at least on a 'doses per patient-year' basis) derive directly or indirectly from such early 'systems biology'-based drug discovery. In the interest of speed and the perceived advantages of mechanistic insight, however, animal models were successively replaced with tissue-level screens (e.g., vascular or tracheal muscle tone), simple cell-based pathway screens (proliferation, cytokine production) and finally with today's ultra-high-throughput screens capable of interrogating individual molecular targets with hundreds of thousands of compounds a day.

[1]Laboratory of Immunology and Vascular Biology, Department of Pathology (5324), Stanford University School of Medicine, Stanford, California 94305-5324, USA. [2]The Veterans Affairs Palo Alto Health Care System, Palo Alto, California 94304, USA. [3]BioSeek Inc., 863-C Mitten Rd., Burlingame, California 94010, USA. Correspondence should be addressed to E.C.B. (ebutcher@stanford.edu).

Published online 6 October 2004; doi:10.1038/nbt1017

Today's 'win-by-numbers' approach is very powerful when applied to known, validated targets (which often means targets of historical drugs), but has led to disappointingly few new drugs when applied to less well biologically understood (e.g., genome-derived) targets. The desire to mine the wealth of the genome has come face to face with the realization that knowing a target is not the same as knowing what the target does, let alone knowing the effects of a chemical inhibitor in diverse disease settings. In fact, despite the enormous investment in genomics and screening technologies over the past 20 years, the cost of new drug discovery continues to rise while approval rates fall[1]. The primary selection of drug targets and candidates has become divorced from the complexity of disease physiology. Reenter systems biology, in modern guise.

The goal of modern systems biology is to understand physiology and disease from the level of molecular pathways, regulatory networks, cells, tissues, organs and ultimately the whole organism. As currently employed, the term 'systems biology' encompasses many different approaches and models for probing and understanding biological complexity, and studies of many organisms from bacteria to man. Much of the academic focus is on developing fundamental computational and informatics tools required to integrate large amounts of reductionist data (global gene expression, proteomic and metabolomic data) into models of regulatory networks and cell behavior. Because biological complexity is an exponential function of the number of system components and the interactions between them, and escalates at each additional level of organization (**Fig. 1**), such efforts are currently limited to simple organisms or to specific minimal pathways (and generally in very specific cell and environmental contexts) in higher organisms[2–4]. Even if our ability to measure molecules and their functional states and interactions were adequate to the task, computational limitations alone would prohibit our understanding of cell and tissue behavior from the molecular level. Thus, methodologies that filter information for relevance, such as biological context and experimental knowledge of cellular and higher level system responses, will be critical for successful understanding of different levels of organization in systems biology research.

This review focuses on recent advances in the practical applications of systems biology to drug discovery. Three principal approaches are discussed (**Fig. 1**): informatic integration of 'omics' data sets (a bottom-up approach); computer modeling of disease or organ system physiology from cell and organ response level information available in the literature (a top-down approach to target selection, clinical indication and clinical trial design); and the use of complex human cell systems themselves to interpret and predict the biological activities of drugs and gene targets (a direct experimental approach to cataloguing complex disease-relevant biological responses). These complementary approaches, which must ultimately be integrated

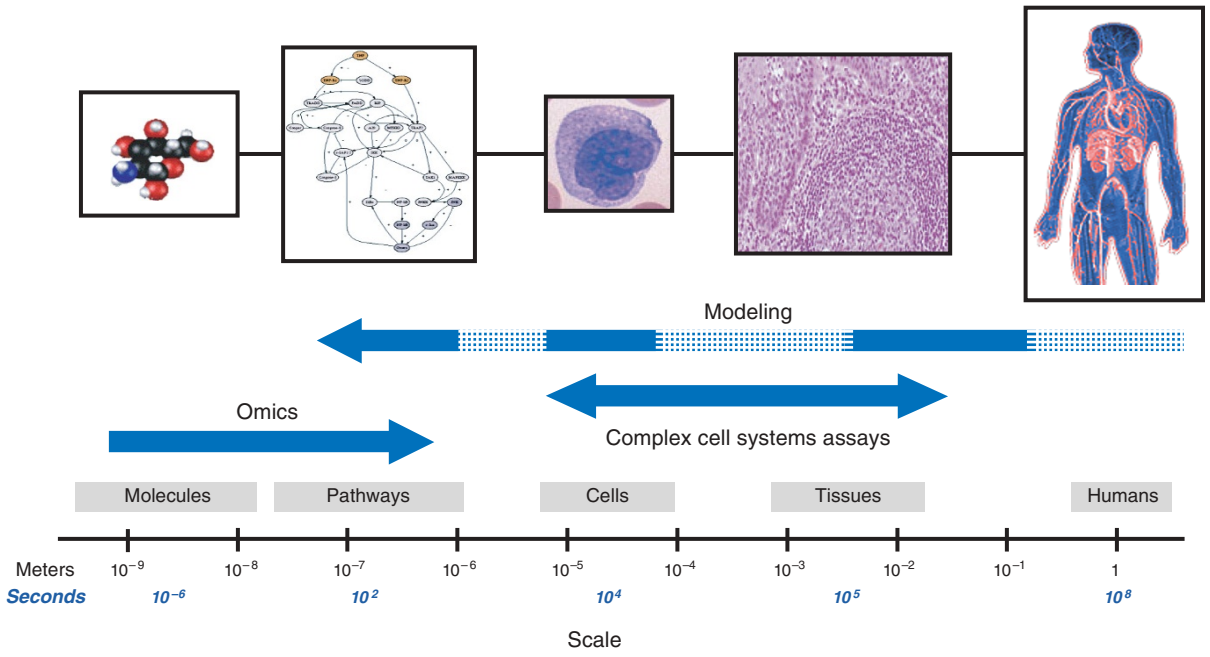

Figure 1 Approaches to systems biology in the pharmaceutical industry. Omics (the bottom-up approach) focuses on the identification and global measurement of molecular components. Modeling (the top-down approach) attempts to form integrative (across scales) models of human physiology and disease, although with current technologies, such modeling focuses on relatively specific questions at particular scales, e.g., at the pathway or organ levels. An intermediate approach, with the potential to bridge the two, is to generate profiling data (e.g., biologically multiplexed activity profiling or BioMAP data) from high-throughput assays designed to incorporate biological complexity at multiple levels: multiple interacting active pathways, multiple intercommunicating cell types and multiple different environments. Such a complex cell systems approach addresses the need for data on cell responses to physiological stimuli and to pharmaceutical agents as an aid to modelers, and also as a practical approach to systems biology at the cell signaling network and cell-cell interaction scales.

in the quest for a hierarchical, molecule-to-systems level understanding of human disease, are already having an impact on the drug discovery process.

Omics: large-scale data generation and mining

It could be argued that a full understanding of the responses of a system requires knowledge of all of its component parts. Omics approaches to systems biology focus on the building blocks of complex systems (genes, proteins and metabolites). These approaches have been adopted wholeheartedly by the drug industry to complement traditional approaches to target identification and validation, for generating hypotheses and for experimental analysis in traditional hypothesis-based methods. For example, omics can be used to ask what genes, proteins or phosphorylation states of proteins are expressed or upregulated in a disease process, leading to the testable hypothesis that the regulated species are important to disease induction or progression (**Table 1**). Integration of genomics, proteomics and metabolite measurements within the context of controlled gene or drug perturbations of complex cell and animal models (and in the context of clinical data) is the basis of systems biology efforts at a number of drug companies, including Eli Lilly (Indianapolis, IN, USA), where they are accelerating the study of complex physiological processes such as bone metabolism[5].

Omics classification of disease states can lead to more efficient targeting or even personalization of therapies by identifying the specific molecular pathways active in particular disease states and in individual patients[6]. Another valuable application of the technology is the identification of surrogate markers for disease detection, or for monitoring of therapies[7,8]. Although omics approaches thus accelerate development of mechanistic hypotheses and clinical insights, a systems-level understanding does not automatically emerge.

Significant efforts are underway to understand key pathway and organism-level responses by relying on the emergent properties of global gene and protein expression data (that is, the properties of the system as a whole that cannot be predicted from the parts). In relatively simple organisms, studies incorporating analysis of time-series genome-wide mRNA expression data, large-scale perturbation analyses and identification of coregulated components, and protein-protein interaction studies have led to new insights into pathway functions and signaling network organization in specific biological processes, such as cell proliferation or the response to metabolic perturbation[9–12]. Although the added levels of complexity in human disease, as well as economic and computational limitations severely limit the utility of omics as a stand-alone approach for systems-level understanding, omics technologies will be important for constructing the 'scaffolds' that help define and limit the possible pathways and connectivities in top-down models of cell-signaling networks[3].

Computer models: from pathways to disease physiology

The goal of modeling in systems biology is to provide a framework for hypothesis generation and prediction based on *in silico* simulation of human disease biology across the multiple distance and time scales of an organism (from molecular reactions to organism homeostasis and disease responses)[2,4]. We are certainly a long way from achieving any general, integrated model of human cell behavior, let alone human organismal biology, but real progress is being made in developing and testing computational and experimental methods for *in silico* systems biology at different scales (**Table 2**). Moreover, we do not need a global

Table 1 Uses for, and challenges of, each systems biology approach: omics, complex cell systems and modeling

	Omics	Complex cell systems[b]	Modeling
Systems level insights	+/–[a]	+	+
Hypothesis generation	+	+	+
Model testing	–	+	–
Target identification/ validation	+	+	+/–
Compound validation	+/–	+	–
Lead optimization/SAR	–	+	–
Disease indication/ trial design	–	+/–	+
Throughput	Slow	Fast	Very slow
Challenges	Quantity of data (curse of dimensionality) Data quality Need for biological context	Availability of cell types Limited modeling of systemic effects	Missing or erroneous data ('garbage in, garbage out') Model validation difficult

[a]Approach cannot address issue, –; approach can address issue, +; approach can address issue under certain conditions, +/–.
[b]e.g., see BIOMAP, **Fig. 3**.

synthesis for modeling and simulation to be useful for basic biological insights and drug development; highly focused, problem-directed models are already having an impact on target validation and clinical development decisions (**Table 1**).

Mathematical and more recently computational models have a rich history in human physiology[4,13–15]. Modeling efforts useful for drug discovery and development must simulate responses at the scale of cell and tissue or organ complexity (that is, the scale at which disease manifests itself). At the same time, a sufficient level of detail must be included such that intervention points accessible to drug discovery are available and can be modulated *in silico* to predict an organ level readout. Thus, a model simulation of heart contractility must incorporate the connection between Na^+/Ca^{2+} exchangers and contractility to be useful to predict the effect of drugs targeting these channels[14]. Difficulty arises in developing models that can effectively integrate the molecular, cellular and organ levels. In addition to pure computational issues, limitations in bottom-up knowledge and in our understanding of pathway and network architecture and interactions, as well as a general lack of standardized knowledge of cell- and tissue-level responses to bioactive stimuli that could be used to validate models (see below) are fundamental, long-term problems that have to be addressed before models integrating complexity at multiple scales can be considered.

A practical approach to address the computational issues is to put in place an organ-level framework and add increasing complexity in a modular format. For example, one can begin with models of inflammation that examine cell-cell communication through cytokine networks and then start replacing the 'black box' cells with simulations of cell behavior (**Table 2**) modeled from network modules (e.g., models of cytoskeleton motility, proliferative or cytokine responses), ultimately replacing 'black box' pathway modules with bottom-up approaches[4].

Entelos (Foster City, CA, USA) has developed complex simulations of disease physiology using a framework of deterministic differential equations based on empirical data in humans[16] (**Table 2**). In these models, internal signaling pathways are not modeled explicitly; cells or even tissues are represented as black boxes that respond to inputs by giving specified outputs that vary with time. Using such an organ level 'disease physiology' framework, Stokes *et al.*[17] have developed a computational model of chronic asthma that incorporates interactions among cells and some of the complexity of their responses to each other and their environment. Model parameters can be modified to reach a particular steady state reference point, for example, the state of chronic asthma (including chronic eosinophilic inflammation, chronic airway obstruction, airway hyperresponsiveness and elevated IgE levels) or the state of exercise-induced airway obstruction. Simulated 'asthmatics' respond as expected to various drugs, including β_2 agonists, glucocorticoids and leukotriene antagonists[17]. Moreover, by simulating an antibody-dependent reduction

Table 2 Examples of computational models relevant to human disease biology

Approach	System	Comments	Reference
Disease physiology	Heart	Review of quantitative models of the heart from genes to physiology	14
	Diabetes	Review of approaches for modeling diabetes	21
	Asthma	Computationally-based mathematical model of chronic asthma predicts patient responses to different therapies	17
		Model of chronic asthma predicts lack of efficacy of IL-5 inhibition	18
Integrative cell models	Cancer	Network model containing 1,000 genes/proteins, 3,000 components predicted effect of specific gene knockdowns	34
	Cardiomyocyte	Approach for linking modules (intermediary metabolism, electrophysiology and mechanics) for developing computational model of cardiomyocytes	15
Pathway models[a]	Multiple	Emergent properties (extended signal duration, threshold behaviors) of signaling in network models	23
	EGFR/MAPK	Review of computational models of EGFR signaling	35
		Ligand concentration change rate versus affinity in downstream outcomes	36
		Describes executable logic model of EGFR network based on rewrite rules	37
	NF-κB	Delay features of signal propagation mechanisms and sensitivity characteristics of signaling components	38
		Reveals role of IκB isoforms in bimodal signal processing characteristics of NF-κB signaling pathway	39

[a]See http://www.cellml.org/examples/repository/ for more examples.

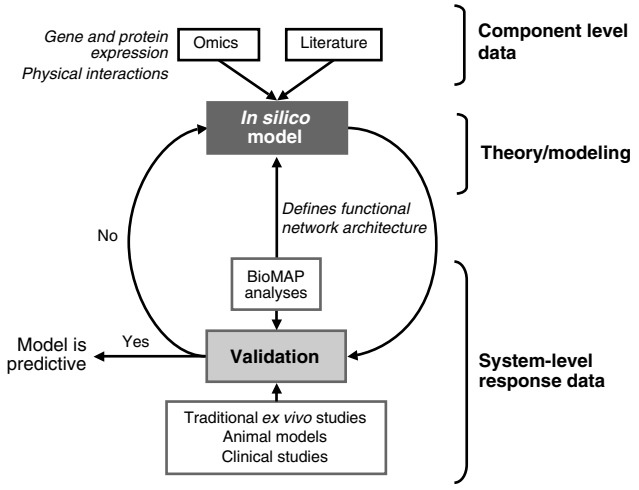

Figure 2 Development cycle of integrated *in silico* models using component level and system response data. Integrated models of disease can be generated using data from the literature as well as protein expression and interaction data sets, potentially informed by predictions of functional network organization and cell responses based ideally on complex human cell-based assays (e.g., see **Fig. 3**). Models are iteratively tested and improved by comparison of predictions with systems (cell, tissue or organism) level responses measured experimentally through traditional assays or from profiles generated from complex, activated human cell mixtures under a set of different environmental conditions. Component level 'omics' data can provide a scaffold, limiting the range of possible models at the molecular level.

in interleukin (IL)-5 protein (a driver of eosinophilia during asthma), this model predicts a decrease in airway eosinophilia but little therapeutic improvement in airway conductance[18], predictions that are consistent with the results of a clinical trial testing a humanized anti-IL-5 antibody in asthmatics[19].

Similar cell- and organ-scale models of glucose metabolism and homeostasis have a long history, evolving from simple relationships between glucose and insulin levels in circulation[20] to more complex models involving integrated multiple tissue responses and their involvement in glucose metabolism[21]. A presentation of Entelos' diabetes 'PhysioLab' at a recent conference (*In Silico* Biology Conference, San Diego, California, USA, June 2–3, 2002; C. Wallwork, personal communication) described how such a computational model has been used in the design of phase 1 trials for an unspecified drug treatment for type 2 diabetes. The results suggested that computational modeling enabled the experimental dosing arms and the number of patients required for the trial to be decreased, thus potentially reducing costs and increasing the probability of clinical success.

More detailed understanding of the systems behavior of intercellular signaling pathways, such as the identification of key nodes or regulatory points in networks or better understanding of crosstalk between pathways, can also help predict drug target effects and their translation to organ and organism level physiology. To this end, a very large number (more than can be fairly cited) of efforts have been focused at the scale of signaling pathways within cells (e.g., see **Table 2**). These models benefit from the large amount of literature data and the promise that omics efforts can provide constraints on the pathways (see previous 'Omics: large-scale data generation and mining' section). As for cell- and organ-level models, simulations of mammalian signaling networks usually rely on time-dependent differential equations and model the pathway in isolation and under very specific (and simple) conditions[3,22]. A next level of detail that enhances the utility of such pathway models is the crosstalk between pathways. Bhalla *et al.*[23] modeled signaling modules and found that combinations of simple modules lead to nonlinear responses or 'emergent properties' of the system. These nonobvious results based on pathway nonlinearity hold promise for identification and prioritization of intervention points within signaling networks.

Interestingly, the architecture of signaling pathways displays significant conservation during evolution, an insight that is being used to help define and understand mammalian cell signaling pathways based on homology with well-defined pathways in lower organisms, and between evolutionarily duplicated pathways in man (e.g., the PathBlast tool[24]). However, although pathway homologies may suggest conservation of key points for chemical intervention in signaling, divergence of pathway functions and regulatory interactions are the norm so that ultimately there can be no substitute for studies in complex human systems.

No matter how successful current attempts at predictive modeling turn out to be, such models raise the challenge of experimental validation (theoretically, only possible with human data) and the cycles of improvement inherent to the modeling effort[3] (**Fig. 2**). From a drug discovery point of view, any of the successes to date could be considered anecdotal and until a given model shows a track record of successful prediction in humans, it will be risky to rely on it for development decisions. For the foreseeable future, modeling predictions will likely be one of many inputs into the decision making process in the pharmaceutical industry.

Using complex cell systems to assay and model biology

Pathway modeling as yet remains too disconnected from systemic disease biology to have a significant impact on drug discovery. Top-down modeling at the cell-to-organ and organism scale shows promise, but is extremely dependent on contextual cell response data. Moreover, to bridge the gap between omics and modeling, we need to collect a different type of cell biology data—data that incorporate the complexity and emergent properties of cell regulatory systems and yet ideally are reproducible and amenable to storing in databases, sharing and quantitative analysis.

At one extreme, responses of human tissues themselves can be probed *ex vivo*, an approach that, even with limitations in terms of availability and reproducibility of human tissues, has proven useful for validating selected compounds and targets[25]. Highly reproducible or even automated approaches to cell biology, however, seem more likely to contribute to the large-scale compound and gene function analyses desired by industry and required as a basis for modeling efforts. Indeed, high-throughput cell-based screening systems, often relying on reporter assays and cell lines, are being used effectively by many companies to identify components of pathways[26], screen for active compounds[27] and even to profile drugs based on their effects on pathway or simple stimulus-response readouts[28,29]. However, these assays are generally designed to isolate individual pathways and to minimize biological complexity and thus neither take advantage of, nor provide insight into, emergent properties of cell systems. This 'systematic biology' focus on simplified pathways is thus to be distinguished from the 'systems biology' focus on complexity and emergent properties.

At the same time, some groups are beginning to appreciate the importance of emergent properties in drug development. For instance, researchers at CombinatoRx (Boston, MA, USA) search for novel combination therapies by taking advantage of two stimuli (phorbol myristate acetate, an activator of the protein kinase C cascade, and ionomycin, a stimulator of Ca^{2+} dependent signaling) that turn on

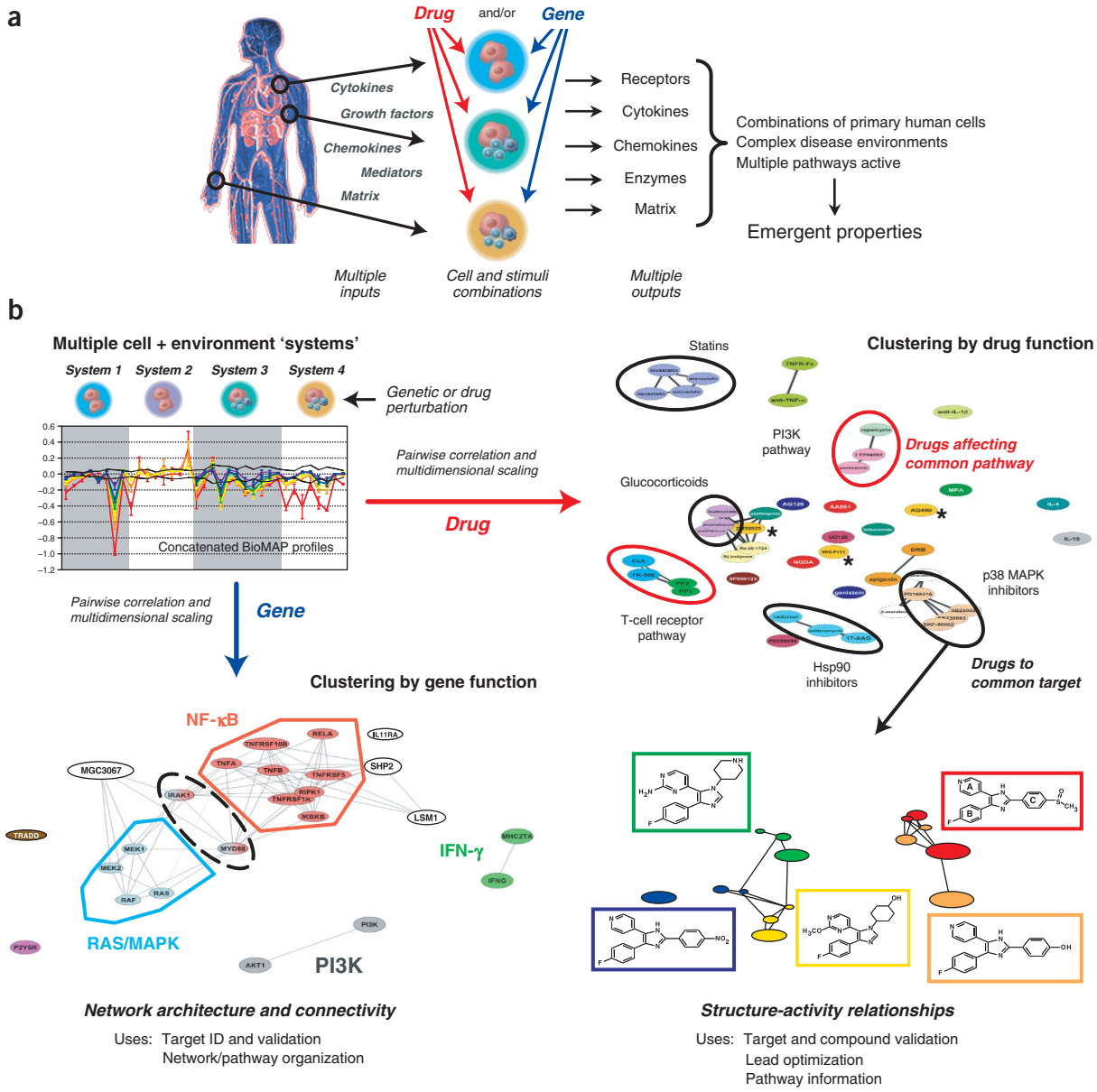

Figure 3 Leveraging complexity in cell systems biology for drug discovery: biologically multiplexed activity profiling (BioMAP) applied to gene function, network architecture and drug activity relationships. (**a**) Primary cells (e.g., endothelial cells and/or blood lymphocytes) are combined and exposed to stimuli (e.g., cytokines, growth factors or chemical mediators) in combinations relevant to the disease biology of interest (e.g., inflammation). Readouts used to measure system responses can be proteins, activated states of proteins, genes or other cellular constituents or properties selected for disease relevance (e.g., cytokines, growth factors, adhesion receptors, which are the ultimate mediators of cellular communication and function in disease) and for responsiveness to environmental and pharmacologic inputs (information content). Perturbations to the parallel systems define the biological activity profiles of interrogating drugs or genes. The combination of multiple cell types and multiple pathways activated elicits complex network regulation and emergent properties that enhance the sensitivity and ability of the systems to discriminate unique drug and gene effects. (**b**) Several complex human cell 'systems' (cells or cell combinations in disease-relevant environments) are interrogated with genes (via overexpression or siRNA) or drugs of interest and the effects on the levels of selected protein readouts are determined, generating a profile that serves as a multisystem signature of the function of the test agent. Statistical measures of profile similarity (*i.e.,* do particular agents induce the same multisystem response?) can be used to cluster genes or drugs by function, and to generate graphical representations of their functional relationships with each other[28,29]. As examples, clustering of profiles induced by gene overexpression (bottom left) reveals key pathway relationships (e.g., Ras/MAPK, phosphatidyl inositol 3-kinase (PI3K), interferon-γ (IFN-γ), and NF-κB-associated clusters) as well as pathway–pathway interactions in signaling networks controlling endothelial cell responses in the context of different inflammatory cytokines[32]. Clustering of drug-induced profiles from inflammatory model systems (comprising activated combinations of endothelial cells and peripheral blood mononuclear cells) detects and discriminates the activities of most known modulators of inflammation as well as a surprising array of other drug targets and pathways, including for example glucocorticoids, cytokine antagonists, and inhibitors of HMG-CoA reductase, calcineurin, inosine monophosphate dehydrogenase, phophodiesterases, nuclear hormone receptors, phosphatidyl inositol 3 kinases, heat shock protein 90, casein kinase 2, janus-activated kinases, and p38 MAPK among others (illustrated in upper right; drugs are colored by mechanistic class)[28,30]. Drugs specific for a common target (circled in black) or for targets in a common pathway (circled in red) cluster together, but compounds having different off target activities are readily detected (e.g., the profiles of three JAK inhibitors with known secondary activities; asterisks). Clustering of activity profiles from lead chemical series can define compound-specific structure-activity relationships for lead optimization (lower right; different analogs are color coded; circle size reflects concentration). In the example shown, BioMAP clustering defines two functional activity classes among structurally related p38 MAPK inhibitors.

multiple pathways in primary cells to search for pairs of compounds that exhibit antagonism (e.g., to tumor necrosis factor (TNF)–α secretion from activated T cells) when combined, but not when used singly[28]. Elsewhere, Rosetta Inpharmatics (Seattle, WA, USA) has measured thousands of output genes in yeast, using the gene response profiles resulting from genetic or chemical (drug) perturbations to determine how genes that effect growth fit into pathways[12] and to reveal the mechanism(s) of action of compounds[29]. These experimental approaches have begun to harness the power of systems biology, but the systems studied remain intentionally simple, focusing on only a few inputs or outputs (CombinatoRx) or a single physiologic state in a model organism (Rosetta). Complexity is a byproduct, not a product of design of these approaches.

Complexity and emergent properties in biology derive from several features: first, complex inputs that stimulate multiple pathways; second, multiple outputs that are integrated network responses to the inputs; third, interactions between multiple cell types; and fourth, multiple contexts and environments for each cell type or combination of cell types. The drug discovery industry has invested billions of dollars in technologies to evaluate outputs, but to incorporate disease-relevant complexity into drug discovery, intentional efforts must also be made to study cells in combination to mimic cell-cell interactions critical to *in vivo* regulatory networks and to assay cells in different complex environmental contexts (in which different combinations of pathways are activated). Parallel context or 'multisystem' analysis is important because proteins and pathways have evolved to integrate inputs and outputs from multiple contexts, so that to understand the effects of a drug (or target), data must be derived from cell responses in multiple environments.

Our group at BioSeek (Burlingame, CA, USA) has developed human cell–based assays that intentionally incorporate complexity at multiple levels, using parallel interrogation of standardized cell 'systems' (cells plus environments) designed to mimic physiological complexity by including one or more primary cell types as well as combinations of cells and active pathways (**Fig. 3a**). Cell systems are engineered to embody disease-relevant responses for biological function analyses, modeling and drug discovery. For example, a panel of just four cell systems (combinations of endothelial cells and blood mononuclear cells in four different complex inflammatory environments) was found to embody complex biology reflecting distinctive contributions of many pharmacologic targets relevant to inflammation[30,31]. Profiles made up of as few as 24–40 protein readouts (including cytokines, chemokines, adhesion receptors and other inflammatory mediators) used to assess the responses of these complex systems are able to discriminate and classify most of the pathways and mechanisms effected by known modulators of inflammation, as well as a surprising array of other drugs and pathways tested[30,31] (**Fig. 3b**). Importantly, the profiles generated from these complex, activated cell mixtures are reproducible, allowing archiving in databases and automated searching and analyses by profile similarity or other characteristics (e.g., effects on key disease-relevant parameters).

This approach, termed biologically multiplexed activity profiling (BioMAP), has been successfully employed in model studies suggesting its applicability to several stages of the drug discovery process (**Table 1**). For target identification and validation, informatics approaches based on the similarity of database-stored multisystem profiles have been shown to rapidly associate gene or drug activities with known (or novel) pathways, and to predict functional pathways and network interactions[32] (**Fig. 3b**). Multisystem profiles induced by gene overexpression in endothelial cells in four different cytokine environments (in essence, multisystem signatures of gene function) automatically clustered into groups that reflected known pathway relationships with surprising fidelity[32]. Moreover, graphical representation of function similarity relationships (**Fig 3b**, lower left panel) point to unique roles for two gene products, MyD88 and IRAK, in mediating interactions between the nuclear factor (NF)-κB and Ras/mitogen-activated protein kinase (MAPK) pathways. MyD88, previously known to signal via NF-κB, was subsequently confirmed in biochemical studies to trigger the MAPK pathway as well, which in turn inhibited NF-κB activation in a negative feedback loop activated by IL-1β but not TNF-α[32]. Clustering multisystem response profiles, in which the systems are designed to capture emergent properties, can thus help define the functional architecture of signaling networks, information important (in conjunction with conventional data sets) for designing and testing computational models.

For compound characterization, the limited data sets, automation and broad functional coverage may make profiles generated from complex, activated cell mixtures an efficient way to screen focused libraries for effects on complex, disease-relevant biology and, more importantly, to prioritize hits from conventional high-throughput screening. In model studies, we have used profiles in four systems to classify hits and leads by their biological activities, to identify compounds with off-target activities (which may be desirable or undesirable), to distinguish 'well-behaved' lead series displaying consistent biological responses and to monitor structure-function relationships as a guide to lead optimization[31] (**Fig. 3b,** lower right panel).

An additional strength of the multisystem approach is that parallel systems can be designed to capture a wide range of elicited (disease-relevant) biological and pathway activities; thus, the effects of drugs or genes can be assessed simultaneously for complex biological responses relevant to many different diseases and can be used to screen for novel therapeutic indications. (This contrasts with most modeling efforts and even animal or clinical trials, which are typically designed to address a single disease target.) Complex cell systems models of inflammation (**Fig. 3**), for example, readily detect the activities of 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) reductase inhibitors (e.g., statins) on inflammatory signaling[30]. This prompts the interesting question of whether inclusion of complex biological systems analyses in the development of statins could have accelerated the discovery of their potent role in autoimmune and inflammatory disorders[33]?

Omics could and certainly should be applied to cell systems designed to incorporate meaningful biological complexity. However, as indicated by studies by our group, highly informative functional signatures for gene and drug effects can be generated using very small numbers (tens) of biologically significant parameters, when these are assayed within several different complex cell and environment combinations. This appears to bear out the prediction that biological complexity encodes useful information about drug and protein function, and suggests that it can be leveraged for 'smarter, faster, cheaper' industrial-scale functional profiling.

From the practical near-term perspective, these approaches present an opportunity to integrate systems biology more efficiently and cost effectively throughout the drug discovery process. From a fundamental perspective, databases of such quantitative human cell biological responses to drugs and gene alterations, under standardized and reproducible conditions designed to embody disease-relevant complexity and capture emergent properties, are likely to be useful in predicting the functional architecture of complex regulatory networks and will provide an essential bridge for integration of omics data into in silico models of cell systems behavior, as well as a testing ground for these models as they develop (**Fig. 2**).

Conclusions

During drug development, million-dollar decisions are (and must be) routinely made using flawed criteria based on incomplete biological knowledge: for example, targets are prioritized because they are upregulated at the gene level in disease (even though many of our best historical targets are not); compounds are selected to be biochemically specific (though many of our most effective drugs are not); animal models are considered essential (although these are known to be poor predictors of clinical success). Better biology, preferably more relevant to human disease and capable of being integrated into the drug discovery process, is sorely needed to inform decision-making. Although the systems biology approaches outlined here are in their infancy, they are already contributing to meaningful drug development decisions by accelerating hypothesis-driven biology, by modeling specific physiologic problems in target validation or clinical physiology and by providing rapid characterization and interpretation of disease-relevant cell and cell system level responses.

Although these approaches are currently being pursued by separate laboratories and companies, it is clear that they are complementary and that ultimately they must be integrated for systems biology to achieve its potential. An analogy can be drawn to the genome project, in which multiple individual efforts contributed technology and informatics approaches that eventually enabled a concerted 'big science' push to sequence the genome. However, whereas the linear output of the genome project was easily standardized and archived, the multidimensional and multivariate nature of biological function and cell biology studies presents an extraordinary informatics and even social challenge, since standardization of experimental design and data are essential before a 'big science' approach to systems biology can be envisioned. Markup languages for gene expression data, emerging ontologies for sharing and integrating different kinds of omic and conventional biological data[4] and the introduction of standardized high-throughput systems biology and associated informatics approaches represent important first steps on this path.

ACKNOWLEDGMENTS

Writing of this review was supported in part by SBIR grants (R44 AI048255 and R43 AI049048) to BioSeek, Inc., and by NIH grants to E.C.B. The authors thank Evangelos Hytopoulos and Ivan Plavec for thoughtful criticism and input.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Published online at http://www.nature.com/naturebiotechnology/

1. DiMasi, J.A., Hansen, R.W. & Grabowski, H.G. The price of innovation: new estimates of drug development costs. *J. Health Econ.* **22**, 151–185 (2003).
2. Ideker, T., Galitski, T. & Hood, L. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372 (2001).
3. Ideker, T. & Lauffenburger, D. Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol.* **21**, 255–262 (2003).
4. Hunter, P.J. & Borg, T.K. Integration from proteins to organs: the Physiome Project. *Nat. Rev. Mol. Cell Biol.* **4**, 237–243 (2003).
5. Kulkarni, N.H. *et al.* Gene expression profiles classify different classes of bone therapies: PTH, Alendronate and SERMs, Poster 307, 31st European Symposium on Calicified Tissue, June 5, 2004, Nice, France; http://www.ectsoc.org/nice2004/abstracts.htm#onl
6. Weston, A.D. & Hood, L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J. Proteome. Res.* **3**, 179–196 (2004).
7. Clish, C.B. *et al.* Integrative biological analysis of the *APOE*3*-leiden transgenic mouse. *Omics* **8**, 3–13 (2004).
8. Kantor, A.B. *et al.* Biomarker discovery by comprehensive phenotyping for autoimmune diseases. *Clin. Immunol.* **111**, 186–195 (2004).
9. Davidson, E.H. *et al.* A genomic regulatory network for development. *Science* **295**, 1669–1678 (2002).
10. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 (2001).
11. Covert, M.W., Knight, E.M., Reed, J.L., Herrgard, M.J. & Palsson, B.O. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**, 92–96 (2004).
12. Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
13. Crampin, E.J. *et al.* Computational physiology and the Physiome Project. *Exp. Physiol* **89**, 1–26 (2004).
14. Noble, D. Modeling the heart—from genes to cells to the whole organ. *Science* **295**, 1678–1682 (2002).
15. Bassingthwaighte, J.B. & Vinnakota, K.C. The computational integrated myocyte: a view into the virtual heart. *Ann. NY Acad. Sci.* **1015**, 391–404 (2004).
16. Musante, C.J., Lewis, A.K. & Hall, K. Small- and large-scale biosimulation applied to drug discovery and development. *Drug Discov. Today* **7**, S192–S196 (2002).
17. Stokes, C.L. *et al.* A computer model of chronic asthma with application to clinical studies: example of treatment of exercise-induced asthma. *J. Allergy. Clin. Immunol.* **107**, 933 (2001).
18. Lewis, A.K. *et al.* The roles of cells and mediators in a computer model of chronic asthma. *Inter. Arch. Allergy Immunol.* **124**, 282–286 (2001).
19. Leckie, M.J. *et al.* Effects of an interleukin-5 blocking monoclonal antibody on eosinophils, airway hyper-responsiveness, and the late asthmatic response. *Lancet* **356**, 2144–2148 (2000).
20. Bergman, R.N., Ider, Y.Z., Bowden, C.R. & Cobelli, C. Quantitative estimation of insulin sensitivity. *Am. J. Physiol.* **236**, E667–E677 (1979).
21. Kansal, A.R. Modeling approaches to type 2 diabetes. *Diabetes Technol. Ther.* **6**, 39–47 (2004).
22. Eungdamrong, N.J. & Iyengar, R. Modeling cell signaling networks. *Biol. Cell* **96**, 355–362 (2004).
23. Bhalla, U.S. & Iyengar, R. Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387 (1999).
24. Kelley, B.P. *et al.* PathBLAST: a tool for alignment of protein interaction networks. *Nucleic Acids Res.* **32**, W83–W88 (2004).
25. Coleman, R.A., Bowen, W.P., Baines, I.A., Woodrooffe, A.J. & Brown, A.M. Use of human tissue in ADME and safety profiling of development candidates. *Drug Discov. Today* **6**, 1116–1126 (2001).
26. Chanda, S.K. *et al.* Genome-scale functional profiling of the mammalian AP-1 signaling pathway. *Proc. Natl. Acad. Sci. USA* **100**, 12153–12158 (2003).
27. Haggarty, S.J., Koeller, K.M., Wong, J.C., Butcher, R.A. & Schreiber, S.L. Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays. *Chem. Biol.* **10**, 383–396 (2003).
28. Borisy, A.A. *et al.* Systematic discovery of multicomponent therapeutics. *Proc. Natl. Acad. Sci. USA* **100**, 7977–7982 (2003).
29. Marton, M.J. *et al.* Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat. Med.* **4**, 1293–1301 (1998).
30. Kunkel, E.J. *et al.* An integrative biology approach for analysis of drug action in models of human vascular inflammation. *FASEB J.* **18**, 1279–1281 (2004).
31. Kunkel, E.J. *et al.* Rapid structure-activity and selectivity analysis of kinase inhibitors by BioMAP analysis in complex human primary cell-based models. *Assay Drug Dev. Technol.* **2**, 431–441 (2004).
32. Plavec, I. *et al.* Method for analyzing signaling networks in complex cellular systems. *Proc. Natl. Acad. Sci. USA* **101**, 1223–1228 (2004).
33. Mach, F. Statins as novel immunomodulators: from cell to potential clinical benefit. *Thromb. Haemost.* **90**, 607–610 (2003).
34. Christopher, R. *et al.* Data-driven computer simulation of human cancer cell. *Ann. NY Acad. Sci.* **1020**, 132–153 (2004).
35. Wiley, H.S., Shvartsman, S.Y. & Lauffenburger, D.A. Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol.* **13**, 43–50 (2003).
36. Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D. & Muller, G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.* **20**, 370–375 (2002).
37. Eker, S. *et al.* Pathway logic: symbolic analysis of biological signaling. *Pac. Symp. Biocomput.* **7**, 400–412 (2002).
38. Cho, K.H., Shin, S.Y., Lee, H.W. & Wolkenhauer, O. Investigations into the analysis and modeling of the TNF alpha-mediated NF-kappa B-signaling pathway. *Genome Res.* **13**, 2413–2422 (2003).
39. Hoffmann, A., Levchenko, A., Scott, M.L. & Baltimore, D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. *Science* **298**, 1241–1245 (2002).