Systems biology, information, disease and drug discovery

Pharmaceutical companies are confronted with an urgent need to increase their lead compound and clinical candidate portfolios and to satisfy market demands for continued innovation and revenue growth. The lack of improved productivity and limited drug pipeline growth is well described and documented. This is in stark contrast to innovative and critical analysis of how to actually alleviate such problems. Recently, Bains (Drug Discovery World 5: 9-18 2004) suggested that improving scientific and management decisions was key to reducing the monetary and time factors associated with launching a successful drug. In this article an emerging new approach, systems biology, is described that may facilitate many of the current attempts to improve the drug discovery and development process. In part, this is achieved by systematically integrating technologies to create a superior output of data and information that enhances our knowledge and understanding of biological function, and chemico-biological interactions, and allows scientists and managers to make better-informed decisions. Systems biology employs an integrated approach to study and understand the function of biological systems and how perturbations such as therapeutic drug administration affect such systems. The biological system can be at the molecular, sub-cellular organelle, cell, tissue, organ or organism level. In this article we attempt to familiarise the reader with biological complexity and systems biology and discuss its potential impact on the drug discovery and development process.

By Professor Stephen Naylor

At present, it is estimated that the total cost of bringing a therapeutic drug to market via the conventional drug discovery and development (DDD) pipeline is $800 million[1]. Furthermore, and predicated on tools and technologies available today, each successful drug will still take, on average, 10-12 years to meander through the stringent and laborious procedures of research, development and federal regulatory oversight before entering the marketplace. Numerous attempts have been made to alleviate the tremendous pecuniary and temporal hindrances associated

with the DDD of new therapeutic drugs. Most have involved technology changes and include the use of combinatorial libraries, high throughput bioassays and the introduction of individual '-omic' platforms. Unfortunately, all such approaches to date have had limited, and even possible detrimental, impact. For example, it has been noted that there has been a 30-year decline in pharmaceutical research and development productivity, despite ~13% annual growth rates expended on biomedical research by both government and industry[2]. Indeed, even the number of late-stage clinical trials has continued to decrease over the past decade[3].

Bains, in a recent thoughtful and provocative article highlighted a series of factors to consider if we are to improve the DDD process[4]. His initial premise focused on the fact that DDD is not a linear process but iterative, and based on his model the cost of bringing a successful drug to market is actually ~$1.15 billion. He also argued that the introduction of new technologies *per se*, would not significantly impact the improvement of the DDD process, but would only "…provide yet another opportunity for just one more experiment". This ultimately translates into increasing the time factor required to bring a drug to market! Bains correctly points out that poor science, technology and medical understanding also contribute significantly to the ballooning cost and time constraints of the process. However, he also makes the salient point that an additional major contributing component is poor management decisions concerning borderline projects. He audaciously announces that "implementing a ruthless success or die policy could half the cost and time to get a drug to market". Scientists are also not spared in this analysis, and Bains indicates that another significant way to cut cost and time is by reducing "repeat" experimental steps at any stage in the DDD process[4].

Decision making and information

Bains makes a compelling case of the need for both managers and scientists to make unambiguous and decisive decisions. However, in order for both groups to make such decisions, the quality, fidelity, accuracy and interpretability of the available data is of paramount importance. Unfortunately, like all of us in the age of the global communication village, managers and scientists are inundated each day with polybytes of data and information. They are ill-equipped to analyse such content and efficiently utilise it in key decision making processes. Most of the data and information remains unfiltered, unprocessed and unused. Our ability to transform

Data ➔ Information ➔ Knowledge

is particularly limited, since we lack many of the appropriate tools.

Historically, the DDD process has suffered from a paucity of information that can be used to make key scientific and management decisions. In part this is due to the technical difficulties associated with obtaining meaningful measurements on biological systems (eg organisms, organs, tissue, cells, organelles or biomolecular entities) under investigation, which results in limited data output. Furthermore, in the past, DDD has been driven by a reductionist approach, which has led to the development of relatively simple biological models, as well as a rudimentary and incomplete understanding of complex biological processes and systems. However, the 1990s forged the 'decade of measurements', which begat numerous high throughput analytical tools and technologies. The consequence of this 'Omics Revolution' has been the development of platforms that now routinely produce copious and substantial, genetic, genomic, transcriptomic, proteomic, functional proteomic and metabolomic datasets[5]. As these datasets have been acquired and analysed our perspective on biological processes (eg homeostasis), appears to be simplistic. For instance, even at the cellular level, simple biochemical pathways appear to be interconnected, modulated, regulated and significant redundancy built into them[6]. Proteins do not normally function as single entities, but act via stoichiometrically-defined complexes that can contain 10-100s of proteins. The formation and disassembly of such complexes are under remarkable control and modulation elements[7]. It would appear that in the biological and life sciences, 'as we have learned more, we appear to understand less!' This has led to a radical rethinking about how we go about gathering biological data and its conversion into information, and ultimately the production of new understanding and knowledge. Such creative thinking has also slowly begun to impact the world of DDD, as practitioners are awash in omic data, large combinatorial library choices, and ultra-high throughput bioassay platform output. How does one go about interpreting and utilising such data and information in making informed decisions?

History and status of systems biology

One such exciting new approach in data acquisition, analysis and interpretation is systems biology, also referred to as pathway, network, integrative or new biology[8,9]. The development of first generation systems biology is an attempt to establish a more integrated and hierarchical paradigm that

Table 1: Organisations developing *in silico* tools and descriptors for general and specific systems modelling and simulation

PROJECT/ORGANISATION	DESCRIPTION
Alliance for Cellular Signaling http://www.cellularsignaling.org	Identification of all the proteins that comprise the various signalling systems, the assessment of time-dependent information flow through the systems in both normal and pathological states, and finally the reduction of the mass of detailed data into a set of interacting theoretical models that describe cellular signalling
SBML – Systems Biology Markup Language Caltech ERATO Kitano Systems Biology project http://www.sbml.org	Project is developing the Systems Biology Markup Language (SBML) for representation and modelling of the information components in the system. Promotes establishment of a common, model-based description language for systems biology simulation software
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 the creation and exchange of biological models. CellML™ is a collaborative effort between Physiome Sciences and the University of Auckland to develop and maintain a family of Physiome Mark-up Languages that also include AnatML™ and FieldML™ for describing anatomic data and the spatial distribution of biological properties
E-Cell Project Keio University http://www.e-cell.org/about/index.htm	E-CELL is a modelling and simulation environment for biochemical and genetic processes. Defines functions of proteins, protein-protein interactions, protein-DNA interactions, regulation of gene expression and other features of cellular metabolism, as a set of reaction rules
Virtual Cell University of Connecticut Health Center http://www.nrcam.uchc.edu	Virtual Cell environment is applicable to mammalian cells and is based on precise measurement of how molecules diffuse to react with target cells. V-Cell defines physiological model of a cell system of interest as an interacting collection of species, structures and reactions
BioSPICE DARPA BioSPICE Berkeley BioSPICE (consortium of Univ California-Berkeley, Lawrence Berkeley Natl Lab, Stanford Research Inst) http://www.biospice.lbl.gov/home.html	BioSPICE is based on a stoichastic biological network simulator originally designed for modelling gene regulation. Has broad ranging capability in developing computational models of intracellular processes

facilitates the creation of new biological pathways and networks at the cellular level. This should provide a framework for understanding the holistic system of genetic, genomic, protein, metabolite and cellular events that are in flux and interdependent. The field is still in its infancy. However, two distinct approaches have evolved, namely computational systems biology[10-12], and more recently data-derived systems biology[6,8,13,14]. The former relies on computational modelling and simulation tools and is also referred to as top-down systems biology. The latter approach predominantly utilises omic and imaging datasets that are mined in a discovery manner for new knowledge

References

1 Tufts University, Tufts Center for the Study of Drug Development. http://csdd.tufts.edu/InfoServices/Publications.asp.2004

2 Booth, B and Zemmel, R. Prospects for productivity. Nat. Rev. Drug Dis. 3: 451-456 (2004).

3 Anonymous. Parexels Pharmaceutical R & D Statistical Sourcebook 2003/2004. page 42. (Parexels International Corporation, Waltham, MA, USA 2003).

4 Bains, W. Failure rates in drug discovery and development: will we ever get any better? Drug Discovery World 5: 9-18 (2004).

5 Hood, L. A personal view of molecular technology and how it has changed biology. J. Protome Res. 1: 399-409 (2002).

6 Ideker, T et al. Integrated genomic and proteomic analysis of a systematically perturbed metabolic network. Science 292: 929-934 (2001).

7 Naylor, S and Kumar, R. Emerging role of mass spectrometry in structural and functional proteomics. Adv. Protein Chem. 265: 217-235 (2003).

8 Morel, N et al. Introduction to systems biology – A new approach to understanding disease and treatment. Mayo Clin. Proc. 79: 651-658 (2004).

9 Editorial, End of the interlude? Nat. Biotech. 22: 1191 (2004).

10 Kitano, H. Foundations of Systems Biology (2001). MIT Press, Cambridge, USA.

11 Kitano, H. Computational systems biology. Nature 420: 206-210 (2002).

12 Nicholson, JK et al. The challenges of modeling mammalian biocomplexity. Nat. Biotech. 22: 1268-1274 (2004).

13 Clish, C et al. Integrative Biological Analysis of the APOE*3 Transgenic Mouse. OMICS. J. Integrative Biol. 8: 3-13 (2004).

Continued on page 27

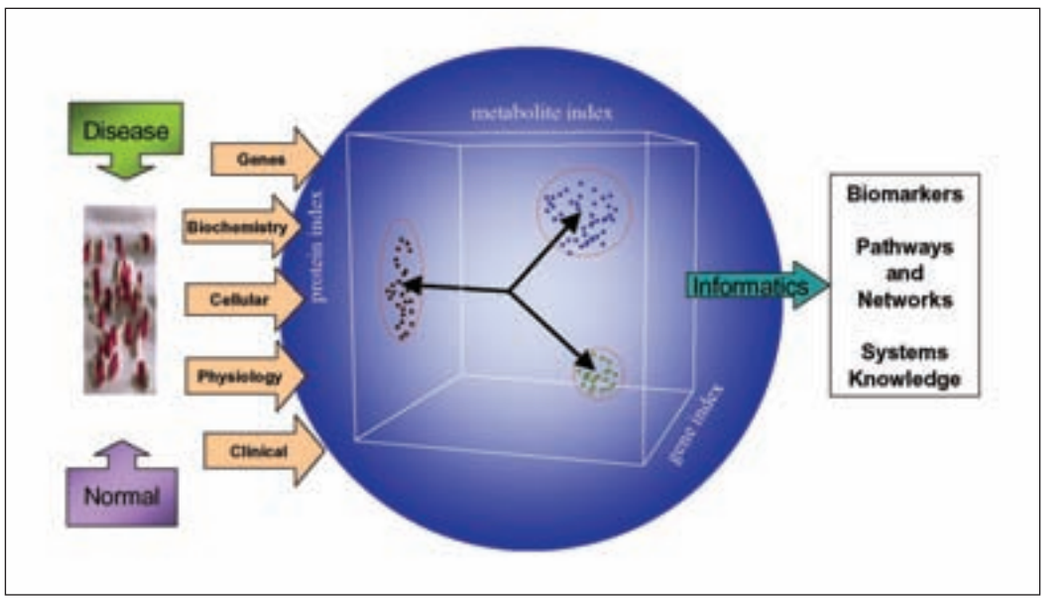

Figure 1
Schematic of conceptual architecture behind systems biology. A series of measurements ranging from genetic/genomic through to clinical are made and a comparison between normal versus perturbed (eg diseased/drug treated/toxin administration) populations is performed. Complex datasets are integrated and a variety of informatic, biostatistical and knowledge assembly tools are used to produce new knowledge and understanding about the perturbed system compared to the normal system. The output can range from molecular and cellular biomarkers to pathways and networks of the system under investigation

using a variety of bioinformatics and knowledge assembly tools, and is often called bottom-up systems biology[15,16].

In the past, systems biology has been predicated on theoretical considerations of complex system analyses. Norbert Weiner introduced mathematical models of complex systems control and communication in the 1940s[17]. However, Ludwig von Bertalanffy wrote in 1928 that "...a [system consists of] a dynamic order of parts and processes standing in mutual interaction... the fundamental task of biology is the discovery of the laws of biochemical systems", and he ultimately went on to develop General Systems Theory[18]. In the 1960s and 70s, Biochemical Systems Theory and Metabolic Control Theory (also now known as Metabolic Control Analysis) attempted to create simple mathematical models of biological systems[19,20]. More recently, Hiroaki Kitano, John Doyle and colleagues at the California Institute of Technology have continued to develop computational systems biology. Kitano defines it as "simulation-based analysis, which tests hypotheses with *in silico* experiments, providing predictions to be tested in *in vitro* and *in vivo* studies"[10,11]. He has gone on to define the four essential elements needed for systems biology to be useful and they include systems structure identification, systems behaviour analysis, systems control and systems design[15]. All these efforts have blossomed into significant computational initiatives, and these are listed in **Table 1**.

The advancement of data-derived systems biology (ie the bottom-up approach) has had to await the advent of the 'omic revolution' and the 'decade of measurements'. This approach has been championed by Lee Hood (Institute for Systems Biology, Seattle, WA, USA), as well as the private company Beyond Genomics (Waltham, MA, USA). Both have attempted to develop a more applied methodology for systems biology analysis[6,8,13]. They have used a variety of omic platforms in concert with sophisticated statistical, bioinformatics and knowledge assembly tools to transform the discovery process by studying, in parallel, complex relationships among genetic, genomic, proteomic and metabolic pathway and networks.

At present systems biology is flourishing, and the phrase has garnered almost scientific pop-culture status. In large part this is due to a number of government initiatives in Europe, as well as major US government agency-led funding efforts (eg NIH-The Roadmap, NSF, FDA and DOE). In addition numerous academic institutions have formed, or are creating systems biology (or integrated biology) programmes/centres/institutes, and they include Harvard, MIT, California Institute of Technology, Duke, University of Michigan, Princeton, Stanford, University of California consortium (Berkeley, San Francisco and Santa Cruz), University of California at San Diego, Southwestern University, University of Washington, University of Manchester/UMIST, University of Leiden/University of Amsterdam, University of Utrecht, Max Plank Institute, Institute for Systems Biology (ISB) and Pacific Northwest Labs.

In the private sector, companies continue their struggle to develop a profitable systems biology business model. However, numerous small

biotechnology companies label themselves as systems biology-like companies and they can be broken down into three different sectors. The first group contains the system biology tool companies and includes Accelrys (accelrys.com; located San Diego, CA; founded 2001); GeneGo (genego.com; New Buffalo, Michigan; 2000); Genpathway (genpathway.com; San Diego, CA; 1999); Ingenuity (Ingenuity.com; Mountain View, CA; 1998); ProSanos (Prosanos.com; La Jolla, CA; 2000); Target Discovery (targetdiscovery.com; Palo Alto; 1999) and 3rd Millenium (3rdmill.com; Waltham, MA; 1996).

A number of companies are attempting to apply systems biology to DDD, either from a top-down approach, or more experimental based (bottom-up) perspective. Companies that provide computational systems biology capability include Entelos (entelos.com; Foster City, CA; 1996); Gene Network International (gene-network.com; Tokyo, Japan; 2001); Gene Network Sciences (gns-biotech.com; Ithaca, NY; 2000); Genomatica (genomatica.com; San Diego, CA; 2000); and Optimata (optimata.com; Ramat-Gan, Israel; 1999). Experimental-based companies include Beyond Genomics (beyondgenomics.com; Waltham, MA; 2000); Bionaut Pharmaceuticals (bionautpharma.com; Cambridge, MA; 2000); BioSeek (bioseekinc.com; Burlingame, CA; 2000); Cellzome (cellzome.com; Heidelberg, Germany; 2000); Icoria (icoria.com; Raleigh-Durham, NC; 1997); and METabolic Explorer (metabolite-explorer.com; St. Beauzire, France; 1999). For a more detailed description of the actual function and business models of each company see references 15 and 21. Finally, as noted above, large pharmaceutical companies continue to evaluate this fledgling field, and for the most part have adopted a wait-and-see attitude. This is due to a variety of factors (see later), and is in stark contrast to the information technology companies such as IBM, Sun and Oracle, who are becoming significant players in the information-rich systems biology domain.

What is systems biology?

In 1964, Supreme Court Justice Potter Stewart attempted to define hard-core pornography by saying: "I shall not today attempt further to define the kinds of material I understand to be embraced… but I know it when I see it." In part this same conundrum of definition applies to systems biology today. A specific definition of systems biology appears to vary as a function of the expertise of the individual pontificating on the issue. This is somewhat intriguing since there does appear to be a general consensus across the academic, biotechnology and pharmaceutical industry sectors that systems biology will indeed 'revolutionise' our understanding of biology, as well as disease onset, mechanisms and progression, and radically improve treatment of such diseases. This optimistic enthusiasm translates across human, animal, plant and microbial biological.

In part the lack of a clear definition for systems biology is due to a number of factors and includes:

- The language of systems biology, which needs to cross numerous scientific disciplines, is not well developed (a glossary of systems biology terms is being developed by University of Stuggart – see http://www.sysbio.de/projects/glossary/index.shtml).
- The practitioners of systems biology come from a multitude of scientific disciplines with their own scientific-centric perspective. They include analytical chemists, mathematicians, statisticians, computational biologists, chemical engineers, biologists (assorted domain expertise), technologists and informaticians.
- Systems biology can either be approached from a top-down or bottom-up perspective. The latter is a more global attempt to identify and understand all component molecules and how they interact, whereas the former is a more focused, targeted analysis pertaining to mechanism. The reality is that most often a hybrid approach is adopted.

In effect, systems biology comprises a myriad of languages and specialists but without a universal translator, and it is little wonder that a universal definition of this burgeoning field has not been defined and widely accepted.

Lee Hood originally defined systems biology as "the study of all the elements in a biological system (all genes, mRNAs, proteins, etc) and their relationships one to another in response to perturbations"[5]. More recently, Hood has broadened the definition to "systems biology represents an analytical approach to the relationship among elements of a system, with the goal of understanding its emergent properties"[22]. Others have argued that it is even more all encompassing than defined by Hood. We have captured assorted definitions in the following amalgam: "Systems biology is the process of interrogating the genetic, genomic, biochemical, cellular, physiological and clinical properties of a system to define and create a system pathway or network that can be used to predicatively model a biological event(s)"[23] and this is portrayed schematically in **Figure 1**.

Continued from page 25

14 Stephanopoulos, G et al. Exploiting biological complexity for strain improvement through systems biology. Nat. Biotech. 22: 1261-1267 (2004).
15 Rubenstein, K. Systems Biology: Key to Unlocking the Value Within the Omics Revolution. (2004) DM&D Publications, Westborough, MA, USA (http://www.drugandmarket.com).
16 Butcher, EC et al. Systems biology in drug discovery. Nat. Biotech. 22: 1253-1259 (2004).
17 Wiener, N. Cybernetics or Control Communication in the Animal and the Machine. (1965), MIT Press, Cambridge, MA, USA.
18 von Bertalanffy, L. General Systems Theory, (1969), Braziller, New York, USA.
19 Viot, EO. Computational Analysis of Biochemical Systems – A Practical Guide for Biochemists and Molecular Biologists. (2000) Cambridge University Press, Cambridge, UK.
20 Heinrich, R and Schuster, S. The Regulation of Cellular Systems. (1996), Chapman & Hall, New York, USA.
21 Mack, GS. Can complexity be commercialized? Nat. Biotech. 22: 1223-1229 (2004).
22 Weston, A.D and Hood, L. Systems biology, proteomics and the future of healthcare: toward predictive, preventive and personalized medicine. J. Proteome Res., 3: 179-196 (2004).
23 Naylor, S and Cavanagh, J. Status of systems biology-does it have a future? Drug Discovery Today-Biosilico 2: 171-174 (2004).
24 Hood, L and Perlmutter, RM. The impact of systems approaches on biological problems in drug discovery. Nat. Biotech. 22: 1215-1217 (2004).
25 Marinissen, MJ and Gutkind, JS. G-Protein-Coupled Receptors and Signaling Networks: Emerging Paradigms. Trends Pharmacol. Sci. 22: 368-376 (2001).

Continued on page 32

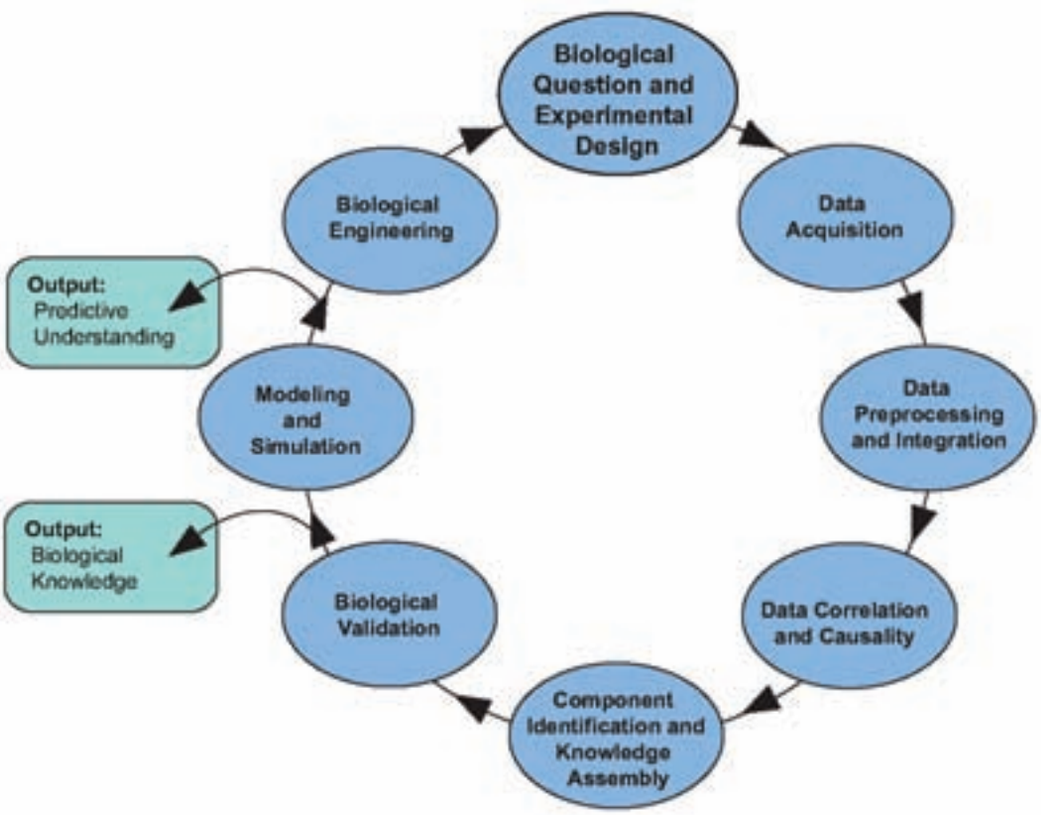

Figure 2
Modules of Systems Biology – it is a process consisting of numerous modules that are interchangeable. The order of the process is determined by the biological question that is being investigated. Output includes new knowledge of a biological system as well as the potential for predictive understanding of that system

Another confounding factor is the question of what constitutes the system under investigation? A 'system' can be something as simple as molecular machinery (eg transcription complex), or an organelle (eg mitochondria), through to a cell, group of cells, tissue, organ or whole organism. Ultimately, the system is defined by the biological question under investigation.

Part of the difficulty in defining systems biology is due to the fact that it is perhaps more appropriate to consider it as a process, rather than a new discipline of biology. The process can be compartmentalised in the form of a series of interconnected modules as shown and summarised in **Figure 2**. The modules comprise:

1. Biological question and experimental design – One needs to select the biological system (ie molecular components, body fluid, organelle, cell type, tissue, organ, organism) to be studied, along with the appropriate hypothesis or discovery driven question to be answered. The necessary controls (eg normal versus diseased), sample histories and outcomes, as well as statistically significant sample numbers to be analysed also need to be determined.

2. Data acquisition – Omic, clinical, physiological and imaging data acquired on a variety of analytical platforms. Datasets are obtained on both control and perturbed (eg diseased, drug or toxin treated, knockout animal) cohorts of samples.

3. Data preprocessing and integration – data files are smoothed, aligned and normalised and ultimately merged into composite files.

4. Data correlation and causality – merged data files are compared and ultimately correlation and causal networks are produced. Tools for data visualisation are also applied at this juncture.

5. Component identification and knowledge assembly – Statistically significant components differing between control versus perturbed are identified. Ultimately, correlation or causal networks are interrogated against all known knowledge using a variety of tools that include simple text mining to semantic web protocols.

6. Biological validation – in order to ensure that the correlation or causal networks have biological relevance, findings must be related back to the biology of the system being investigated. Strategies such as RNA interference, high throughput cellular bioassays and knock-in or knockout organisms can all be employed to provide a biofocusing of data back to the relevance of the biological question originally posed.

7. Modelling and simulation – data and correlation networks can be used as a framework for further modelling and simulation studies of the biological processes under investigation.
8. Biological engineering – once a system has been modelled and understood, one can re-engineer the pathway or network to produce a better outcome, eg disease resistant crop, or identify optimal therapeutic agent.
9. Output: Biological knowledge – relevant biological knowledge is produced to answer in part, or completely, the original question posed.
10. Output: Predictive understanding – based on modelling and simulation studies of the biological system under investigation it should be possible to now predict the outcome of specific changes in the pathway or network.

The specific process flow and the connectivity of the modules are determined by the biological question posed. For example a discovery driven question may require the modular flow described above and shown in **Figure 2**. However, a hypothesis focused question may necessitate a restructuring of modules, such that, for example, knowledge assembly is carried out prior to 'Data acquisition'. In the case of computational approaches, module 7 (Modelling and simulation) is the most likely place to commence. In any event, the modular process of systems biology allows one to capture and portray the staggering complexity and interconnectivity of molecular, cellular, as well as organism events that occur on the microsecond-hour-day-month time-scale.

Systems biology and disease biology discovery

Despite the medical advances made to date, numerous disease states can only be explained by complex, multi-molecular interactions rather than by alteration of a single gene, gene product, or metabolite. Thus, in order to garner a more complete and relevant understanding of disease, one must obtain a comprehensive perspective of the biological system thereby uncovering the interdependent and dynamic pathway, network and cellular events that undergo change as a function of disease predisposition, onset and progression. Hence as noted recently by Hood and Perlmutter[24], describing all the elements of the system and defining the appropriate biological networks, (where disease reflects the operation of perturbed networks), and comparing normal versus diseased networks allows critical nodes to be identified. If such nodal points can be reconfigured back towards the norm, this constitutes a treatment of the disease. This concept is encapsulated in **Figure 3**.

Very recently the first ever systems biology analysis of a mammalian disease model was reported[13]. The apolipoprotein E3-Leiden transgenic mice (APOE*3-Leiden) is a widely used animal model for atherosclerosis. It is a knock-in mouse that expresses a human mutant apolipoprotein E that is normally associated with familial dysbetalipoproteinemia in humans. These mice are highly susceptible to diet-induced hyperlipoproteinemia and atherosclerosis due to diminished hepatic LDL receptor recognition, but when fed a normal chow diet they display only mild type I (macrophage

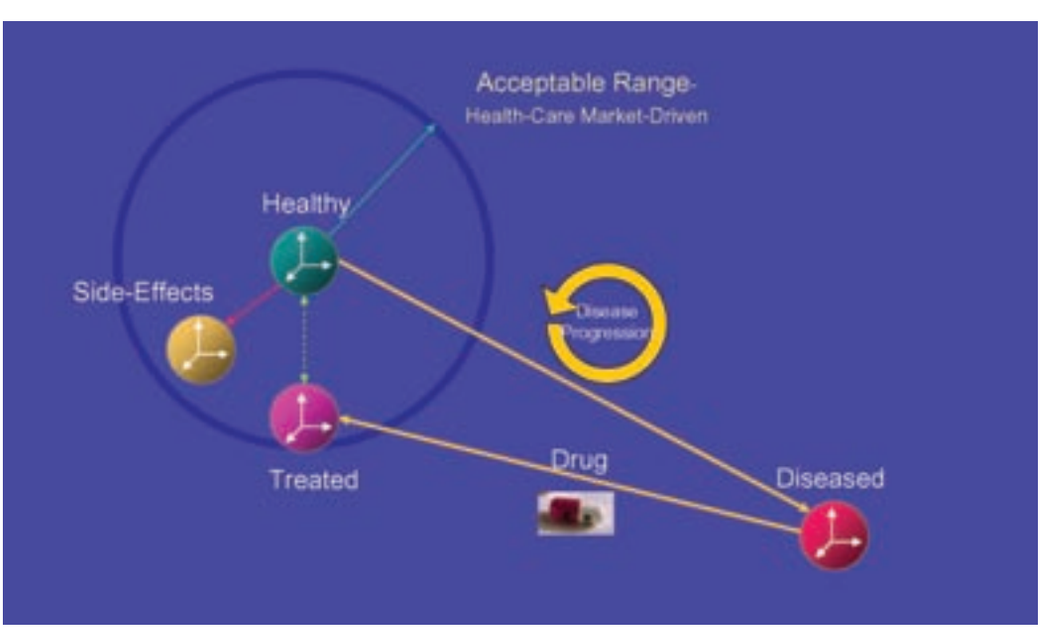

Figure 3
Systems biology paradigm of determining individual health. A more complete, information-rich composite at the molecular and cellular levels is obtained. If an individual's molecular or cellular co-ordinates move outside of the norm, this constitutes a disease state. Treatment with therapeutic agents is designed to move the co-ordinates of the individual back to the original values, if possible. (Courtesy of Dr Eric Neumann, formerly at Beyond Genomics, now at Sanofi-Aventis)

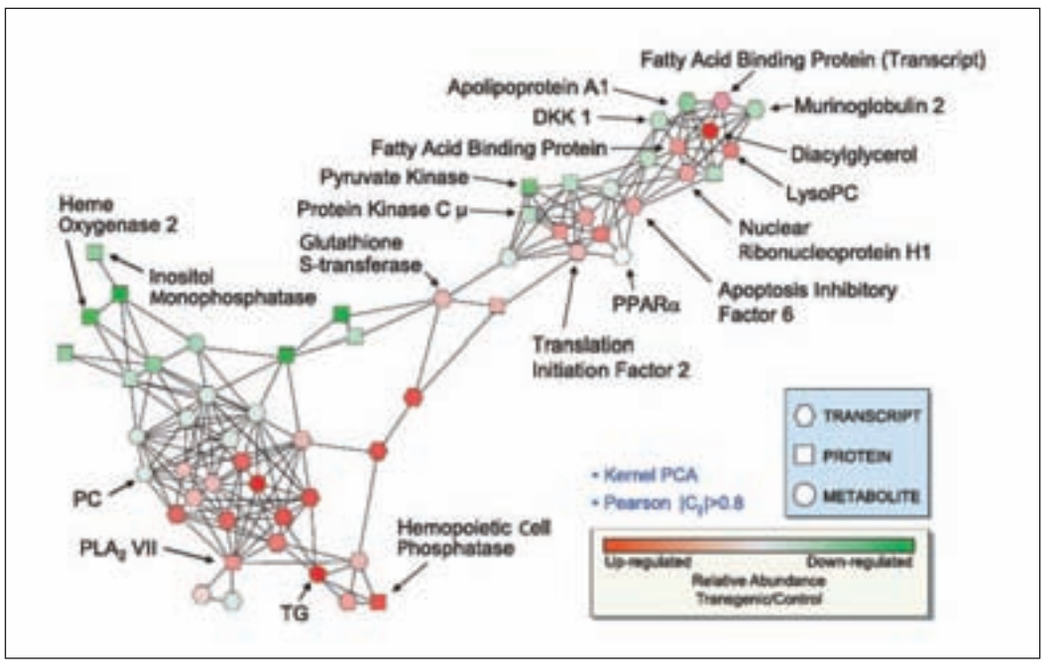

Figure 4
An example of a systems biology analysis output of a diseased animal model. The APOE*3-Leiden mouse has been used extensively as a model for atherosclerosis. Ten transgenic versus isogenic control cohorts were used in this study. The figure shows the Correlation Network of select expressed genes, proteins and lipids. The shading inside the box indicates the relative amount in the transgenic APOE*3-Leiden mice compared to wild type controls (red = higher level, green = lower level) and a black line connecting two entities indicates a high level of correlation (a Pearson correlation coefficient of 0.8 was used as a cut-off). (Published with Permission OMICS: J Integrative Biology13)

foam cells) and II (fatty streaks with intracellular lipid accumulation) lesions at nine months.

Liver tissue was obtained from individual isogenic wild type mice and APOE*3-Leiden mice that were fed a normal chow diet and sacrificed at nine weeks of age. In order to carry out a systems biology analysis, individual transcriptomic, proteomic and metabolomic data sets for each mouse liver (three files per mouse) were merged into single data files. The resulting 20 integrated datafiles, now containing transcript, protein and metabolite information were then aligned and normalised in order to carry out a variety of statistical analyses. Finally, in the hope of discovering hitherto unknown relationships that might exist among measured genes, proteins and metabolites, a correlation network analysis was carried out. A partial correlation network containing approximately 10% of the overall dataset acquired in the ApoE*3 systems biology analysis is shown in **Figure 4**. This correlation network represents identified proteins and metabolites, as well as genes that either encode the identified proteins or those whose expression were significantly different in wild type versus APOE*3-Leiden mice. The network analysis shows a high degree of correlation among ApoA-I gene, L-FABP (both gene and protein expression), and lipids, such as diacylglycerol and LysoPCs as these components appear as a cluster within the network. Components of this cluster also show a high correlation to proteins involved in metabolism (eg pyruvate kinase) as well as signal transduction (protein kinase C).

After obtaining the correlation network, it was subjected to a variety of knowledge assembly inquiries, including text mining. Known biological connections and pathway constituents were detected, which had previously been reported in fatty acid metabolism studies. However it is interesting to note that less than 10% of the correlations corresponded to known biological connectivities reported previously. Clearly, such an approach affords a powerful new tool in discovery disease biology. This study demonstrates the utility of an integrated approach for characterisation of a highly complex system. By generating high content analytical output and comparing principle component factors derived from composite data sets the authors were able to rapidly elucidate identities and the relative abundances of major lipoprotein metabolism mediators that define the ApoE*3-Leiden phenotype compared to the isogenic control.

Systems biology and drug discovery

The potential utility of systems biology to affect our understanding of disease mechanism and progression can be extended to understand the mechanism of action of therapeutic agents for existing treatments of disease (see **Figure 3**). For example signalling events within or between cells are not restricted to linear pathways, but are well recognised to be components of complex and dynamic networks[25]. For this reason the true mechanism of action for many therapeutics has eluded investigators. Deciphering the complexities of inter- and

Continued from page 27

26 Hopkins, AL and Groom, CR. The Druggable Genome. Nat Rev. Drug Discov. 1: 727-732 (2002).
27 Terret, NK et al. 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: 1819-1825 (1996).
28 Capdeville, R et al. Glivec (STI571, Imatinib), a Rationally Developed, Targeted Anticancer Drug. Nat. Rev. Drug Discov. 1: 493-499 (2002).
29 Alm, E and Arkin, AP. Biological networks. Curr. Opin. Struct. Biol.13: 193-204 (2003).

intracellular signalling relationships is a mission uniquely feasible through systems biology since the fundamental strength of this approach is its ability to analyse an entire system comprehensively, as demonstrated in **Figure 4** and described above. A comparative analysis of treated and untreated diseased samples (ie drug perturbation studies) will facilitate an understanding of signalling networks associated with drug treatment. Specifically, through a comprehensive analysis of the biologic system after the treatment with small molecules we may define molecular consequences affected by these molecules. By the strategic selection of a panel of perturbing agents (ie drugs) for parallel studies, one may be able to discriminate between cellular changes associated with therapeutic benefits and those associated with side effects of various agents – information relevant for development and regulatory purposes. The discovery of novel sites of interaction within the signalling pathway where molecular targets are free of the connection to side effects provides the prospect for therapeutics with enhanced efficacy, reduced side-effects and improved therapeutic indices, and diminished onset to efficacy[26].

One intriguing prospect, as discussed above in some detail, is that systems biology is poised to make an impact on therapy through its ability to define causal mechanisms of disease expeditiously. Causative targets may be well-characterised proteins that have previously lacked a recognised connection to the disease in question (eg the discovery of the connection of phosphodiesterase type 5 enzyme to erectile dysfunction[27]). They may also be novel molecular targets such as endogenous (wild type) or mutant proteins that have been formerly uncharacterised and therefore do not have an established connection to disease (eg the identification of the leukemia-associated protein encoded by the BCR-ABL gene28). Either form of target can be uncovered through a systems biology analysis of the differences between normal and diseased samples.

Finally, it is important to return to a consideration of how systems biology can aid in the decision-making processes of both the scientists and managers involved in the DDD process (see above). However, before evaluating such matters it is interesting to note that the Bain's model indicates an average total cost of bringing a successful drug to market of ~$1.15 billion[4]. He actually breaks down that number into component parts, estimating the "cost per successful project" for Discovery ($609 million; 52.96% of total cost), Preclinical ($39 million; 3.39%), Phase I ($56 million; 4.87%), Phase II ($341 million; 29.65%) and Phase III ($105 million, 9.13%). The discovery process consumes the largest dollar amount in "cost per successful project" at $609 million. This is an intriguing dollar value since discovery biology and chemistry/pharmacology are rife with difficult, subjective go-no go decision points. In the case of clinical trials, matters are compounded by having to deal with human complexity, but many of the decision points are tightly regulated by government oversight. Could Bain's assertion that better scientific and managerial decision-making would cut in half the time and cost of bringing a drug to market actually be correct?

In order to test such a hypothesis, scientists and managers have to have the appropriate tools available in order to make informed decisions. The premise presented here is that systems biology and the output from such analyses affords tools to effectively link data to information and ultimately new knowledge of the system under investigation. These tools and knowledge, when handled correctly, should allow scientists and managers to make better, more informed decisions. This is summarised in **Figure 5**. The advent of the systems biology process has to date had little impact on the pharmaceutical industry. In part this is due to the embryonic state of systems biology. However, the implementation of this multidisciplinary process is not a trivial task and requires expertise and technologies from many different arenas. Amalgamation of these skills in the vast confines of pharmaceutical companies is a difficult undertaking, particularly when considerable silo-like mentalities exist. However, could a rallying cry to the industry be: "It's the process, stupid" (with due deference to James Carville!)?

Future perspectives

It has been suggested that "Systems Biology is ... really a metaphor that points to an interdisciplinary, integrative approach to biology"[15]. However, many believe that systems biology is more than a metaphor; it actually reflects a true paradigm shift of thinking in how biology is investigated, and reflects the arrival of 'big biology'. Some members of the biological scientific community now understand what their physicist colleagues have known for many years, that nature is a tough taskmaster. Irrespective of whether you study her at the subatomic level or the cellular level, she only imparts information and understanding with reluctance. We can only conquer the biological complexity of cellular and organism events with concerted effort. This requires the interdisciplinary skills of polyglots and polymaths as well as new departments/centres/institutes of excellence, in order to provide core

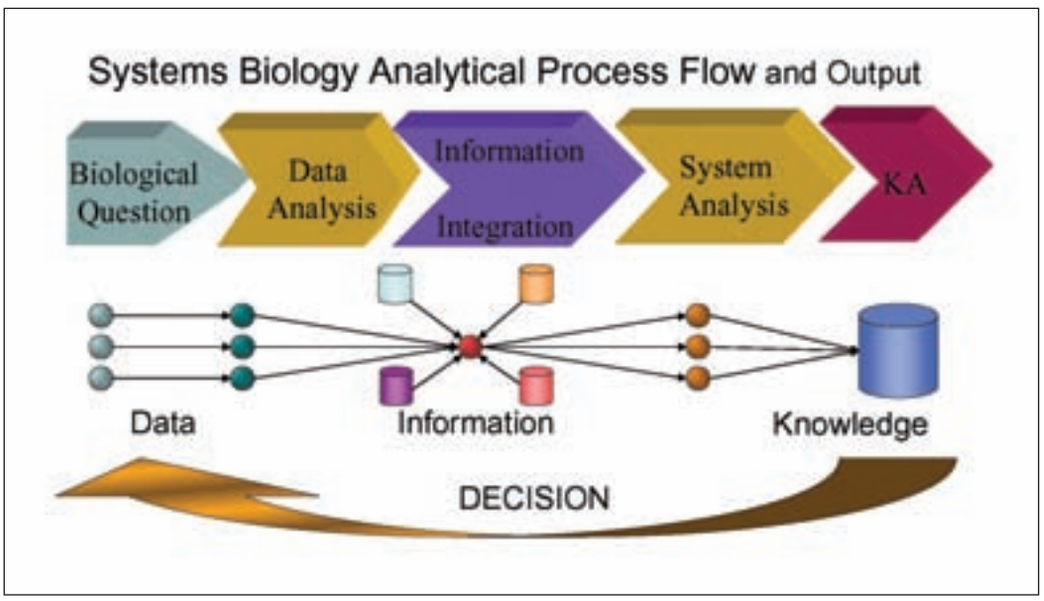

Figure 5
Systems biology process flow and how it may aid in scientific and managerial decision-making for drug discovery and development personnel. (KA is knowledge assembly)

competency in technologies, information and computer sciences and knowledge.

The impact of systems biology on current individual scientific interactions, including DDD, is not trivial. A number of issues need to be considered and discussed for the fledgling efforts presently under way to be successful. Arkin and others have argued that the practice of biology within this new paradigm requires open source sharing[29]. This is similar in concept to the open source computing movement epitomised by companies such as Linux. Data sharing and co-operative science need to be openly encouraged, something that many current scientists find difficult in the present scientific environment. In particular the silo mentality of many large pharmaceutical company groups is largely the antithesis of such open source data sharing.

The advent of systems biology should allow the efficient transformation of substantial, integrated data sets ➔ information ➔ knowledge. In order to achieve such a task, innovative new technologies, data management and integration tools, as well as knowledge assembly capability are needed to build on the first generation approaches already available. The potential impact of systems biology can provide a new dynamic to invigorate biotech and pharmaceutical companies, as well as many other sectors associated with the health and life sciences. Systems biology has the potential of helping pharmaceutical scientists and managers make better informed decisions predicted on a more complete understanding of the problem at hand. Is this the new beginning that the pharmaceutical industry needs, and does systems biology really help solve the decision-making issues so decisively enunciated by Bains[4]? Only time will tell! DDW

Professor Stephen Naylor *is currently Adjunct Professor of Genetics and Genomics at Boston University of Medicine (Boston, MA, USA), as well as a Visiting Faculty Member in the Division of Biological Engineering at MIT (Cambridge, MA, USA) and a Faculty Member of the Computational Systems Biology Initiative (CSBi) also at MIT. He is the former Chief Technology Officer, and Senior Vice-President for Research at Beyond Genomics where, in conjunction with his colleagues, he built the world's first integrated systems biology platform, consisting of both analytical, bioinformatic and knowledge assembly capability. Previously he was the founding Director of the Biomedical Mass Spectrometry and Functional Proteomics Centre at the Mayo Clinic. In addition he was Professor of Biochemistry and Molecular Biology and Professor of Molecular Pharmacology and Experimental Therapeutics. He was also Adjunct Professor of Clinical Pharmacology, as well as Biomedical Engineering (Molecular Biophysics) at the Mayo Foundation. Stephen received his PhD from Cambridge University (UK) in biological mass spectrometry, completed post doctoral work at MIT (USA) and served as Associate Director of Mass Spectrometry at the MRC Toxicology Institute in London. Professor Naylor also serves as a consultant to a number of analytical, pharmaceutical and biotechnology companies, has published more than 225 research papers, has filed a number of patents and made more than 600 presentations at seminars worldwide.*