Current Trends
in Science

PLATINUM JUBILEE SPECIAL

Systems biology and medicine

DHIRAJ KUMAR and KANURY V S RAO

Immunology Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110 067, India.
e-mail: dhirajkrverma@gmail.com

1. Introduction

The traditional approach to the analytical study of natural sciences, including biology and medicine, has often been inclined to take the reductionist route [1]. The roots of such reductionism stem from the general assumption that nature is governed by simple laws. As a result, the most obvious strategy for understanding complex natural systems in general was to reduce the system into smaller, simpler, and therefore more tractable, units [2]. Given the myriad of molecules/units that perform diverse biological functions, researchers were faced with the onerous task of characterizing and detailing the properties of as many of these individual molecules as possible. These investigations have proven to be extremely fruitful, giving us today a detailed cataloguing of the various components that constitute biological systems. In addition, they have also provided insights into the function of a large number of these components. The quantum of information generated through the reductionist approach over the last several decades has not only helped in understanding the biology of the human body in some detail, it has also allowed us to identify the aberrations that occur under conditions of various disease conditions [3]. Implicit in the philosophy of reductionism is the view point that deviations in the behavior of only one, or at best a few, molecule(s) is (are) responsible for a given pathological condition, and the development of any therapeutic strategy should be aimed at modulating/inhibiting this deviant behavior. The range of successes achieved through this route has helped to transform medical science, and helped to combat and control a variety of physiological conditions including pathogenic diseases.

With the turn of the century, however, we witnessed revolutionary changes that altered the very core of how biological research was performed, and the nature of questions being asked. This was primarily spurred by the advent of new high-throughput experimental tools that permitted us to simultaneously monitor the effects of any condition on virtually all the genes, or their protein products, encoded within the human genome. This more global view of biological processes also led to the gradual realization that perturbations in biological systems are never restricted to a single molecule, or even a small group of molecules. Rather, any given perturbation (e.g. mutation, depletion, etc.) influenced the steady state of a large number of the components of the system. Such results led to the gradual realization that biomolecules do not act independently. Rather, they interact with other biomolecules to form tightly integrated networks, and biological responses derive from the behavior of such networks. For example, our own group has previously shown that when a component (a participating molecule) in a signaling network is perturbed, the constituent nodes of the entire network are affected, and it is this change in network properties that defines alterations in processing and transmission of receptor-activated intracellular signals [4,5]. Similarly, in a recent report, Cohen *et al* have examined the distribution of over a thousand different tagged molecules inside the cell, in response to a drug [6]. Their results revealed that virtually all of these molecules showed altered behavior in

Keywords. Systems biology; biological networks; drug targets; drug development; bio-markers; host-pathogen interaction.

their expression, localization, distribution when monitored at the level of both kinetic and spatial resolution. Another example that corroborates the 'one target-full network perturbation' concept is cumulatively represented by studies involving microarray experiments. Here, specific perturbation conditions such as targeted-siRNA mediated knockdown, drug treatment or infection by a pathogen have all been shown to lead to dramatic alterations in the gene expression profile of the target cell [7–10]. This new altered state is the direct result of shifts in the dynamic equilibrium between the host components and the perturbed target such that the ripples of this shift alter the dynamics of the system at the global scale. A direct inference from such type of experiments was that, alterations in the properties of a biological system cannot be attributed to a single component, and therefore specific targets could not be singled out for targeting by medicines [2].

2. Limitations of the current approaches in medicine

The science of medicine has particularly adhered to the reductionist approach ever since its inception. The overriding assumption was that any alterations in the physiological state was the result of malfunction, or altered function, of the single most important component and, therefore, if that particular component could be targeted and brought back to normalcy it would achieve reversion of the altered physiological condition to its normal state [1]. Undoubtedly, this approach has proved successful in controlling and treating various diseases and health conditions. However, it leaves little room for any contextual information to be implemented. The importance of context has now become a central theme in biological research. It can be easily understood by considering a relatively simple example. Two individuals with different genetic backgrounds, one being normal and the other harboring an autoimmune disorder, are likely to be treated with identical drug regimes for a common bacterial or viral infection. Such homogeneous treatment of a genetically heterogeneous population will result in significant differences in the efficacy of such generically employed treatment regimes. The recent advancements in our understanding of biological systems – paradoxically, largely through reductionist exercises – has now forced the realization that every individual is not the same, and that a more personalized approach would be needed even though phenotypic symptoms/health conditions may be shared. The hitherto traditional approaches had emphasized focus on the 'one deciding factor' – i.e. 'the invading pathogen' or the 'malfunctioning component' – with the goal of specifically targeting it in order to restore the normal physiological state. In this process, however, one was oblivious of the implications of these individual targets on the overall regulatory network of the system. Undesirable side effects frequently accompany such therapeutic regiments, which are often misconstrued as either non-specific, or off-target effects of the treatment. It is important to emphasize here that labeling such side effects as 'off-target' effects, at least in many instances, could be misleading. Such categorizations may, in fact, highlight the limitations of the current approach of reductionism in medicine.

Within a network, where the various components are intricately inter-related and maintained in a dynamic relationship, targeting even a single component would influence the overall dynamics of the network. Although the targeted molecule may represent a relevant target for a desired state, the consequent effect on other molecules in the network must not be overlooked. The altered dynamical properties caused due to organizational modulation and network topologies may exercise small but accumulative effects over a longer period of time, resulting in a physiological state that is very different from the normal as well as the diseased state under observation. The implication of such a situation can be easily estimated by considering the number of drugs/medicines across the countries that have been retracted/banned from being prescribed or sold due to their undesirable side effects. Some examples of these are presented in table 1.

The diversity of unrelated effects elicited by these drugs is obvious in table 1. Thus, for example, an anti-diarrhoeal drug could result in side effects leading to cancers, or, an anti-worm agent could eventually cause damage to the nerves. If we look deeper into the etiology of side effects, the networked organization of biological system would emerge as the major causative factor for such side effects. Alternatively, the strategy through which the drug was discovered may also be the culprit. Since the approaches used for its discovery and validation were reductionist, it may not appropriately account for the intricacies of the complex biological organization. The apparent success of these medicines is attributable to the fact that the therapy targets an important factor for the physiological state being treated. The networked organization of the component would however also ensure unintentional perturbation of several other components, which otherwise were functioning normally, thereby leading to undesirable physiological outputs commonly referred to as side effects. We will discuss later in this article the importance of considering the network organization, and design

Table 1. List of drugs banned in India due to their side effects.

Generic name		Use	Reason for ban	Brand name(s)
1.	Analgin	Pain-killer	Bone-marrow depression	Novalgin, Baralgan
2.	Cisapride	Acidity, constipation	Irregular heart beat	Ciza, Syspride
3.	Droperidol	Anti-depressant	Irregular heart beat	Droperol
4.	Furazolidone	Anti-diarrhoeal	Cancer	Furoxone, Lomofen*
5.	Nimesulide	Pain-killer, fever	Liver failure	Nise, Nimulid
6.	Nitrofurazone	Anti-bacterial cream	Cancer	Furacin, Emfurazone
7.	Phenolphthalein	Laxative	Cancer	Jetomisol-P*
8.	Phenylpropanolamine	Cold & cough	Stroke	D'Cold*, Vicks Action 500*
9.	Oxyphenbutazone	NSAID	Bone marrow depression	Sioril
10.	Piperazine	Anti-worms	Nerve damage	Piperazine, Helmazan*
11.	Quiniodochlor	Anti-diarrhoeal	Damage to sight	Enteroquinol

*Denotes it is a combination product.
Analgin, Furazolidone and Nitrofurazone are banned for use even in animals in the United States.
Analgin is banned in Nepal, Vietnam and Nigeria (Reference: MIMS INDIA, September, 2005).

principles of the biological networks, while arriving at any new drug or drug targets.

Yet another limitation of current approaches to drug discovery, especially in the case of pathogenic diseases, is that they frequently induce resistance to the drug by the targeted pathogen [11]. Here again the organizational differences between the pathogen and the host play a decisive role. Thus, the rate of accumulation of advantageous mutations is several-fold higher in the case of pathogens (owing to their relatively simpler genome organization and short multiplication times) as compared to the host. For example, the current regime of therapeutics for most of the pathogenic bacterial diseases relies heavily on select molecules that target one or more metabolic/biochemical process of the pathogen. Most of them target vital processes of the pathogen such as protein biosynthesis (e.g. translation, by inhibiting 30S or 50S ribosomal subunits), nucleic acid metabolism (DNA replication, repair or transcription), or membrane biosynthesis (polymerization of peptidoglycans or other constituents of bacterial/fungal cell wall or membrane). Owing to high multiplication rates, most of the pathogens tend to develop resistance to many of the drugs used, with the drug then serving as a positive selection pressure for the advantageous mutations. Again as we shall see later, an understanding of cross talk between host network and pathogen network could potentially provide us with ways to identify drug targets and drugs that would ensure a foolproof strategy to eliminate the pathogen, without inducing the development of resistance. An analysis involving global set of relationships between protein targets of all drugs and all disease-gene products in the human protein–protein interaction network connected most drugs into a highly interlinked giant component, with strong local clustering of drugs of similar types according to Anatomical Therapeutic Chemical Classification [12]. The harsh reality unraveled by the topological analyses of the resulting network was the identification of overabundance of 'follow-on' drugs, i.e., new drugs that target already targeted proteins. By including drugs currently under investigation, they identified a welcome trend of shifting towards more functionally diverse targets thereby improving the efficacy and longevity of new drugs [12].

An integral part of the revolution in medical science was due to advancements in diagnostic medicine. Diagnostics is truly considered as the first step towards treatment and cure. As expected, however, here again the reins of reductionism have restricted the development of foolproof diagnostic methodologies. The four cornerstones of diagnostic medicines are anatomy, physiology, pathology and psychology. Diverse tools have been developed, and countless markers have been identified that can correlate with the clinical symptoms. The vast majority of these tools, however, only provide an incomplete characterization of the actual cause or condition of the individual. There could be several reasons behind our current inability to acquire biomarkers that specifically pin-point the cause of the state, as well as provide directions to the treatment to be followed. The most prominent of these is the absence of unique markers that specify a given physiological condition. For example, an elevated level of liver enzymes could result from any of the following situations:

- side effect of medication, such as certain non-steroidal anti-inflammatory drugs (NSAIDs), cholesterol medications, antibiotics, or anti-seizure drugs.
- drinking alcohol
- obesity

- diabetes
- elevated triglycerides
- infections such as viral hepatitis and mononucleosis
- autoimmune disorders of the liver and bile ducts, such as autoimmune hepatitis, primary sclerosing cholangitis or primary biliary cirrhosis
- Metabolic liver disease, such as hemochromatosis or Wilson's disease
- Excessive use of certain herbal supplements, such as kava, comfrey, pennyroyal or skullcap.
- Gallstones
- Tumors of the liver, pancreas or bile ducts.

The above list clearly indicates how imprecise and approximate such tests can be. At the very least, such a scenario highlights the limitations of empirical analyses and emphasizes the need for a more systematic methodology for use in diagnostics. At the other end of the spectrum are conditions where several infectious agents escape detection by the currently available methodologies. Here again one could argue that our excessive focus on causative agents/molecules limits the window of opportunities that could potentially be explored. It does not require much of an insight to ask for a radical change in our approach towards diagnostics, bringing in more global information to more appropriately balance cause, effect, and context of a given clinical condition. The current interest in high-throughput biology provides ample scope to monitor alterations between the individual network components at a global level. Wide spectra of information gathered through high-throughput experiments can then be put together to identify a set of patterns that correlates clinically with the symptoms and therefore would serve as foolproof system of diagnosis. The added advantage of such approach would be that one could then probably link diagnostics with tailor-made therapeutic strategies, as well as exploit it for predictive medicine.

3. Why systems biology?

Increase in knowledge always throws up more unknown questions, ensuring that the cycle of new discovery, new knowledge, and new questions is constantly maintained. This aspect is especially exemplified by the human genome sequencing project, the anticipated implications of this project, and the subsequent venturing into addressing multi-parametric questions through systems level approaches. The rough draft of the entire human genetic code was hailed as the most significant map ever produced by human kind [13,14]. This, in turn, spurred the anticipation that the field of medicine would be revolutionized over the coming decades, giving new tests and drugs for previously untreatable diseases. The development of personalized medicines, and even predictive and preventive medicine, did not seem to be a distant dream [14]. However, as we started getting sequence information for the human genome, it became increasingly apparent that genomics alone would not provide answers/solutions to multitude of tasks that were originally thought to be resolvable. Rather, the findings have only revealed that the extent of complexity in biological systems far exceeded the initial anticipations. Nonetheless, the most exciting aspect of the sequencing project was the emergence of an entirely new perspective of looking at biological systems as a whole rather than through the traditional reductionist approach. This perspective also additionally confirmed that genomics alone does not constitute the sole driver for revolutionizing medicine but that radically new concepts were also required. A most telling discovery of the human genome project was that of how many genes humans have. A simple prokaryotic organism like *E. coli* has about 4000 genes, while single celled simple eukaryotes like *Saccharomyces cerevesiae* have roughly about 6000 genes. Given the complexity of the human system it was expected that the total number of genes would roughly be around 100,000. However analysis of the completed genome sequence has put the number finally at less than 19,000. In other words, it just required three times more genes to evolve from a simple unicellular yeast to the complex multi-cellular humans. In spite of this, however, the diversity of functions performed by organs, cells and cellular components of humans are clearly several orders of magnitude greater than any of those in simple organisms. Evidently then, individual protein components in human cells are assigned with multiple tasks, and these components are tightly regulated in order to perform right task at right time and in the right subcellular space. Context, therefore, becomes preeminent at the level of biological organization as the precise task to be performed by the cellular components is defined by the environmental demands placed on the cell. Thus in addition to enlisting molecular components of a system, it is also now important to understand how those components are connected to and regulate each other, as well as the nature of emergent properties that the network of bio-molecules can display. The context-elicited behavior, dynamics, kinetics, and the relative stoichiometry of the cellular components have become some of the most important issues that await resolution. Understanding the topological and functional organization of these gene products in the system at temporally distinct

periods, and the dynamics thereof, is now central to obtaining a deeper insight into biological phenomena. An important development in this regard is that the ability to address such questions is now no more restricted to the realm of the experimental biologist. On the contrary, a more global inter disciplinary collaboration is needed to acquire, manage, analyze, and interpret the large-scale data sets being routinely generated. Data sets comprising gene expression profiles, proteome profiles, phospho-proteome profiles and, later on, interactomes are being continuously produced as a result of the new-found interest in the systems approach. The recent explosion in technologies yielding 'high-throughput' information, therefore, is enforcing a paradigm-shift in the way we approach the study of biological systems and their operational mechanisms [3]. At one level, the ability to simultaneously derive information on thousands of biomolecules has called for new methodologies for analysis of such 'data-rich' results. And as is the case with any dialectical process, even a preliminary analysis of such kind of data has brought to us a new level of realization. Much like an individual in a society, no individual bio-molecule behaves in a truly independent fashion with a discrete biological function. Most biological phenomena are now recognized to represent the integrated outcome of a complex interplay between the individual constituents of the machinery. In other words, a biological response is the net output of oscillations in the properties of a large number of molecules, all acting in an inter-dependent fashion to form a highly connected network.

As is evident from the sections discussed above, a major challenge of contemporary biology is to embark on an integrated theoretical and experimental program to map out, understand and model in quantifiable terms, the topological and dynamic properties of the various networks that control the behavior of the cell. Rapid developments in complex network theory in the past few years have facilitated advances towards uncovering the organizing principles that govern the formation and evolution of various complex technological and social networks [15]. It has also yielded insights into the shared architectural features between biological and other complex networks such as the internet and society. This universality suggests that a common set of laws may govern most complex networks in nature. As a result, information from other complex but non-biological systems can well be employed to dissect the intricate and interrelated relationships that underlie cellular functions [3,16]. The accruing results are shaping the realization that, notwithstanding the importance of individual molecules, cellular function is a contextual attribute of strict and quantifiable patterns of interactions between the myriad of cellular constituents. The realization that biological responses stem from the aggregate properties of underlying molecular networks is now inspiring a more 'holistic' approach for study, and a separate interdisciplinary field of 'systems biology' has now begun to evolve.

4. Systems biology: The evolution of a new discipline

True to its newly found focus, the scientific community is yet to reach a consensus on the definition of systems biology. However, despite its youthful status, there is a growing perception that systems biology may be synonymous with the analysis of large networks that describe the properties of entire genomes, or the properties of the proteome and the corresponding interactome, the comprehensive mapping and functional integration of metabolic pathways, or the combination of all of these systems at different scales of biological organization [17].

The tantalizing appeal of systems biology is its focus on comprehensiveness, although it would perhaps be unwise to look upon it as an exclusive approach. Biological systems are not just large they are also organizationally complex. This complexity is expressed through properties such as dynamics, regulation, and adaptation. However, many of these features tend to express in subtle ways and are often difficult to capture or analyze at the level of all-encompassing systems. Nonetheless, these features constitute the central properties that significantly govern the life and responsiveness of cells and organisms. Therefore one cannot ignore the investigation of their specific roles and functions. Because nonlinear systems such as biological systems are inherently complex, a more practical route that is frequently taken is to rigorously analyze representative examples that can still help to unravel successful patterns of design and operation. In similarity with many of the non-biological networks studied so far (e.g., the worldwide web), it is generally assumed that much of biological organization is hierarchical and modular. Therefore, insight into a variety of smaller systems, or modules (e.g., signaling network, gene regulatory network, etc.), can help to create a foundation on which a deeper understanding of the functionality of large-scale integrated systems can eventually be achieved [17].

Operationally, systems biology represents the study of the behavior of complex biological organization and processes in terms of the molecular constituents. As characterized by Kirschner [18], it is built on molecular biology in its special concern for information transfer, on physiology

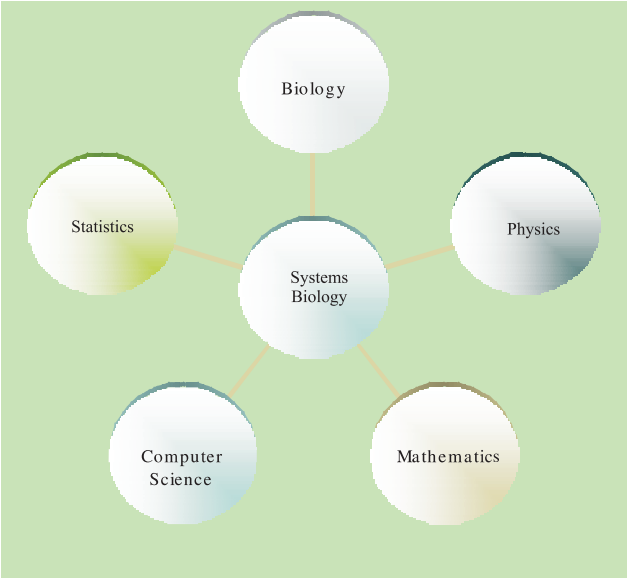

Figure 1. The evolution of the academic network that is being spurred by developments in systems biology.

for its special concern with adaptive states of the cell and organism, on developmental biology for the importance of defining a succession of physiological states in that process, and on evolutionary biology and ecology for the appreciation that all aspects of the organism are products of selection, a selection we rarely understand at molecular level. Systems biology attempts all of this through quantitative measurement, modeling, reconstruction, and theory. Therefore systems biology is not a branch of physics. Rather it applies physics to understand how biology generates variation. Such an imperative to create and understand variation never exists in the physical world [18]. It is important to note that the systems biology perspective brings into focus several additional dimensions of biological processes that have hitherto escaped our attention. This includes the need for a comprehensive and quantitative analysis of molecular functions and interactions in a manner that accommodates its spatial, temporal, and sequential dimensions. The consequent fall-out of this requirement is the need to bring together researchers from diverse disciplines viz. molecular biology, biochemistry, proteomics and genomics, physiology, physics, mathematics, statistics and computer sciences among many others (figure 1). It would be pertinent at this stage to engage in a brief discussion about networks, their properties, their inherent complexity, and also the general approaches for the study of such complex systems.

4.1 *Biological networks*

The post genomic era has converged to an important idea in biology where cells can be viewed as a complex network of interacting proteins, nucleic acids and other metabolites/bio-molecules. Similar complex networks have also been used to describe social networks, internet as well as transport systems [15]. In spite of the significant differences between these systems, they do share common features in terms of network topology and organizing principles. But networks are inherently difficult to understand, as illustrated by the following description of its general features: Networks display structural complexity in that the wiring diagram in general is very intricate. This intricate relationship becomes much more complex due to evolution over time, implicating topological reorganization of the network over a period of time. In addition, various links within a network could have varying degree of weightages and directionality depending on the stoichiometry and relative abundance and/or activity. Biological networks display dynamical complexity. Therefore nodes could be linked with each other in a nonlinear dynamical fashion. Additionally, all these added features can influence each other, further adding to the hierarchical complexity in the organization [15,19].

A key property of these complex networks is their ability to withstand diverse perturbations without collapsing. This property, also known as robustness, is a hallmark of complex networks and constitutes a common feature for all the networks irrespective of the system to which they belong. Robustness defines the system's ability to retain function despite perturbations. On the flip side, robustness is also coupled with fragility towards non-trivial rearrangements of the connections between the system's internal parts [20]. This paradoxical balance between robustness and sensitivity in complex networks deserves a brief discussion.

4.2 *Network structure and topology: Robustness versus sensitivity*

Biological systems are continuously influenced by the environment and need to respond to them in order to survive and grow. Many specific signals to biological systems, for example, cells of an organism are conferred through the environment and cells need to identify those signals and respond accordingly (i.e. in a context-specific manner). The problem arises because such signals may often just represent normal fluctuations in the environment. Cells are therefore left with a difficult task of distinguishing between environmental noise and specific signals. In addition to distinguishing between signal and noise, cells also need to ensure that they mount an appropriate phenotypic response depending upon the nature of the

signal (e.g., proliferate in response to a mitogenic stimulus). While such seemingly difficult tasks are routinely performed with ease, the larger question of maintaining plasticity at the level of the response while ensuring that the network stays robust emerges as an interesting intellectual conundrum. A typical example of the dual requirements for sensitivity and robustness is the immune system. Here, although a high sensitivity to antigens is necessary for inducing an immune response, oversensitization frequently results in allergic reactions or autoimmune diseases. It is not yet clear how the complementary attributes of sensitivity and robustness are incorporated into a single system, although this clearly appears as an inbuilt property in several natural systems. Are there design principles that enable one to build in both sensitivity to the specific signal and robustness in the face of environmental noises [21]?

Recent efforts to understand the behavior of complex biological, social and technological systems have concentrated on the topology of the network that is formed by the connections (links) between the individual components (nodes). In particular, the distribution of links among the nodes is thought to encapsulate information on the capabilities of the system. Most studies of networks in biological and social systems have, implicitly or explicitly, taken the average number of links in the shortest path between an arbitrary pair of nodes as a defining feature, with fewer links being considered to be optimal. The shortest-path viewpoint directly derives from the 'small-worlds' property that was originally discovered for social networks [15,19]. However, the functional utility of this property is still a matter of debate. Most of the complex networks show scale-free topology, meaning that the number of nodes having k links (the most elementary characteristic of a node is its degree or connectivity, k, which defines the number of links that each node has to other nodes) decreases as k to some power for large k. While such features can contribute to the robustness of a system they, however, do not explain its sensitivity attributes [21]. However, contrary to the belief of scale-free organization as a means to achieve robustness for the system, there are also alternate suggestions that instead of highly connected nodes, sparsely connected ones are shouldered with larger functional responsibility in stabilizing and proper functioning of the network and therefore the system [22]. This latter concept is further supported by a report demonstrating that when the costs of complexity are taken into account, robustness correlates with a parsimonious network structure that is sparsely connected and not unnecessarily complex; and that selection will favor sparse networks when network topology is free to evolve [23].

Functional characteristics of some complex systems and their network topologies are better understood in terms of the system's need to respond sensitively to external change by switching from one mode of behavior to another [21]. This is particularly evident in the case of signal transduction and metabolic networks, both of which guide the response of cellular systems to external stimuli, or to modulations in the availability of resources. Responsiveness implies the capacity to adjust even to small environmental changes. In contrast, robustness entails a lack of sensitivity to environmental variations, retaining the same behavior even when subject to large stimuli. Both of these contradictory properties are obligatory for effective maintenance in an ever-changing environment. The topology of the system yields direct information about the nature of the response, in terms of how response (or, information?) is propagated through the network of connections. The complementary requirements of robustness and sensitivity remains an issue that awaits clear resolution and, perhaps, also a re-visitation of the current dogma that scale-free network topologies are robust to failure and sensitive to attack [24–27]. However, generalizing – as is often done – that the former is an advantage, whereas the latter is a disadvantage, may be misleading in some contexts. This is because both characteristics can be advantageous if the sensitivity allows the system to respond effectively to environmental changes.

5. The human body as a complex biological network

The structural and biochemical organization of the human body is inherently complex. Based on various properties and functions that they perform, the diverse constituent systems are named as immune, skeletal, nervous, digestive, endocrine and respiratory systems. Each system comprises diverse organs and organ systems that in turn is made up of different tissues. Tissues are made up of cells and cells are made of thousands of biomolecules including proteins, nucleic acids, carbohydrates and other small metabolites. At each level of organization, individual components of a system are heavily interconnected and cross regulated. At the same time, however, they are also influenced by components from other systems. Thus the concept of scaling in biological systems (figure 2) is an important one, as also is the existence of hierarchical modular network as the organizational principle.

Figure 2 shows an attempt to represent the hierarchical organization and corresponding complexity at respective scales of a human body.

Traditionally cells are considered as the basic unit of life although they are not at the bottom of the hierarchal organizational scheme of a multicellular organism. Cells are built of hundreds of thousands of small and large molecules, majority of them organic molecules, polymers and many inorganic ions. Most of the organic molecules can be broken down into their constituents in terms of elements like carbon, hydrogen, oxygen, nitrogen and sulfur. There does not seem to be much of a complexity involved at the scale of these molecules in the sense that organizational principles at this scale in terms of various covalent bonds and how they can be made and broken is well understood. As we move up in the scale, the interaction among variety of these organic inorganic molecules together constitute subcellular organelles like plasma membrane, mitochondria and nucleus. How are the simpler constituents assembled together to form these subcellular structures? For example, do we know what makes nucleic acids, proteins and lipids together to assemble into nucleus and perform the functions assigned to it? Similarly how organelles assemble together to form a cell and various kinds of cells communicate with each other to form various tissues, organs and organ system that eventually gives rise to the whole body. This self assembly of individual components into higher level organizational scales, with difficult mechanism to be attributed to and the properties displayed by them is called emergent behavior of complex systems [28]. The emergent behavior of complex systems implies that individual components of a system exhibit an interdependent dynamics that eventually define the properties of the system.

Therefore, it also implies that it is relatively very simple to break down a system into its components and study their roles individually than taking all the components and assemble them into a higher order organization. The interactions among the components and resulting dynamics are integral to every genre of organizational hierarchy from molecular to cellular to tissue, organ or whole organism. Not only that, similar principles also govern the complex organization and properties of ecological and social networks.

For a deeper insight into the issues discussed above, we consider an example of a physiological process that results in elimination of rogue cells from the body of a multicellular organism. This process is called apoptosis and is one of the most crucial physiological events that governs and maintains homeostasis in the body. An aberrant apoptotic response is attributed for various physiological disorders and diseases including cancer. The molecular components, dynamic regulation among them as well as underlying mechanisms have been

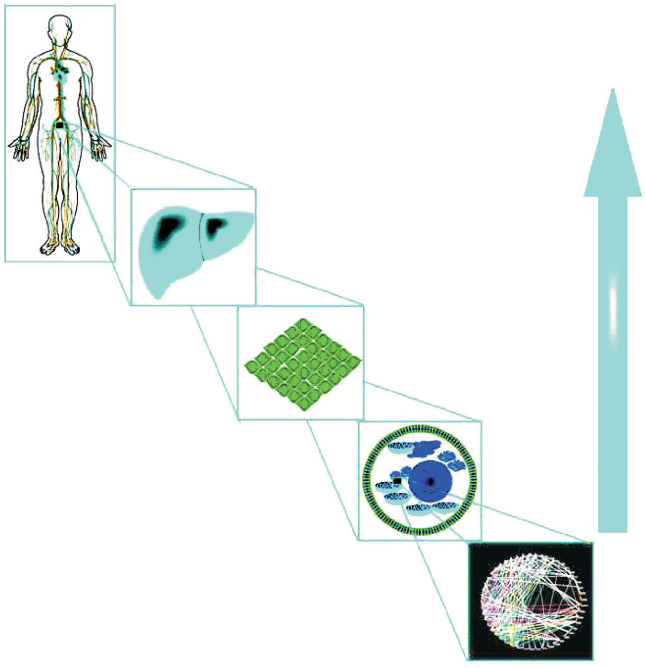

Figure 2. Scaling in biological systems. Organizational hierarchy of the human body: An organism is reduced to its organ, tissue, cellular and subcellular level. At each level of organization, complex interplay among the components establishes and regulates the dynamic property of the system. There are also across-the-scale regulatory relationships, like small molecules (cytokines, hormones) affecting cells, tissues or the whole organ. At every scale biomolecules are rendered with the responsibility of communicating among and across the scales.

the focus of intense research in the past and these continue to be one of the most studied phenomena of the biological system. Expectedly the information generated about this process over decades has followed typical reductionist approaches and thereby, now with the large amount of information available to us it has become extremely complex to understand the regulatory events leading to apoptosis. The scheme in figure 3 represents only the terminal regulatory module for apoptosis without including the background regulatory events. Bax molecules need to assemble together for the most critical and committing step for apoptosis; that is the release of cytochrome C from the mitochondrial membrane. However, Bax is not permitted to do so as it remains sequestered by another family of molecules called Bcl2. Bcl2 can in turn be engaged by a pro-apoptotic molecule called 'Bad' thereby freeing Bax for multimerization and cytochrome release. Bad can be prevented from binding to Bcl2 by phosphorylation at many of its serine or threonine residues by a variety of cellular kinases including MAP kinases and pro-survival kinase Akt [29–32]. These kinases are part of a much more complex module of intracellular signaling network that is responsible for regulating almost every

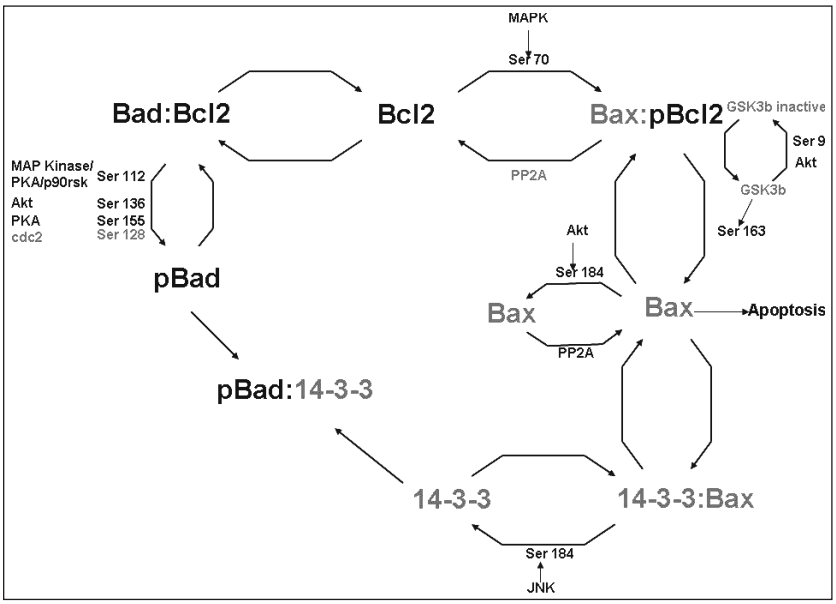

Figure 3. Regulation of biological process and physiological process of apoptosis and complexity of its terminal regulatory module. Majority of signaling events (phosphorylation and de-phosphorylation) eventually regulate the ability of multimerization of Bax in order to regulate apoptosis.

individual function of the cell. Even understanding the dynamical output from this relatively simple schematic is above and beyond the scope and/or ability of the traditional biologist. Here the subject becomes inter-disciplinary, requiring inputs from biologist as well as some mathematician or a physicist to assist in understanding the whole process. Needless to mention then that what would happen when one begins to consider the larger regulatory module involved in regulating this physiological process. It would most likely become impossible to understand without help from experts in other disciplines such as mathematics, physics, statistics and computation. Such collaboration is truly the hallmark of systems biology and it also underscores the current need to understand the biology at a much higher level of resolution.

6. The dynamics of host-pathogen interaction

As we have discussed so far, biological systems exhibit a dynamic balance between sensitivity and robustness to a wide range of external and internal perturbations. Nearly any kind of perturbation can be accommodated by these systems with certain qualitative and quantitative adjustments of their components. When the perturbation is abiotic in nature, the entire accommodative modulations rest with the host system. It, however, becomes significantly more complicated when the perturbing agent is biotic, since both systems are able to exhibit dynamic regulation. A classical example of the latter is establishment of successful infection of host cells by a pathogen. Successful establishment of host-pathogen interaction becomes extremely important in the case of intracellular pathogens. One such example is *Mycobacterium tuberculosis*, which is viewed as one among the most successful of human pathogens [33]. To keep the scope of the present article focused we will only discuss the ensuing host-pathogen interaction in the case of *Mycobacterium tuberculosis* infection although the principles discussed should be broadly applicable to almost any pathogens. Following infection, *Mycobacterium tuberculosis* resides inside the phagocytic cells, the immune cells responsible for clearing all the invaders inside the host. The dynamics that ensues between the pathogen and the host has now started to emerge though still very little is known in this regard [34–41]. The net output of this dynamical relationship however is qualitatively known and can vary between the extremes of total clearance of the bacteria, to the development of active disease, or the establishment of an asymptomatic infection where bacteria can latently reside inside the host cells for varying periods before being reactivated under specific circumstances. What happens when two biological objects, both with their own set of components and respective dynamics, come together to establish a mutual relationship? Certainly the ripples of one system could be felt in the other one and vice versa, both of them reaching to a slightly different level of intra-system equilibrium that permits coexistence of both the systems. This slightly altered state of both the host as well as the pathogen

could be derived from qualitative and quantitative alteration of the respective system components (genes, proteins or metabolites), and the resultant communication between the two. The directionality of the shift in equilibrium would then decide the eventual output of the process. In view of this, it becomes critical to understand the 'qualitative and quantitative' adjustments in both the systems in order to devise strategies for combating tuberculosis. Therefore, characterizing bio-molecular components and their respective dynamics for both the systems become important. Two critical components of both the systems are the proteins (proteome) and the lipids (lipidome). Alteration in the expression pattern and/or stability of several proteins and their specific post-translational modifications has been reported under various infectious/pathogenic conditions. Many of these proteins are indeed shown to be important for successful pathogenesis. However such studies have tended to provide only a partial image of the overall communication process between host and the pathogen. In contrast, a global analysis of the protein-protein interactome of host system, under both infected and uninfected conditions, would provide an invaluable insight on 'qualitative and quantitative' alterations of the host system when perturbed by a pathogen. A global network analysis of essentiality, homeostasis and robustness thereafter would also equip us with novel tools to specifically target contextually critical interactions so as to hamper pathogen growth and survival within the host cell. As has been shown in figure 4(A), an intracellular pathogen acquires the status of a major hub inside the host network, interacting and influencing simultaneously several of the constituent nodes of the host system. It would be interesting to see how vital these interactions are to the survival of the pathogen. Some preliminary results from our own experiments in case of *Mycobacterium tuberculosis* infection and published reports from others in case of HIV, West Nile virus, Dengue virus and *Plasmodium* and suggests many of these host factors could indeed be vital for the ability of the pathogen to survive and grow within the host system [42–46]. Moreover, many proteins from pathogen might be capable of interacting with and altering the functionality of several host proteins. Identification of such cross-species molecular interactome would be ideal for novel drug target discovery.

Similarly, several metabolites between the two species could be cross-recognized. A detailed comparison of the 'metabolome' of the host cell under infected versus non-infected conditions could provide several interesting insights. One broad class of metabolites widely studied and implicated to be critically important for mycobacteria are lipids

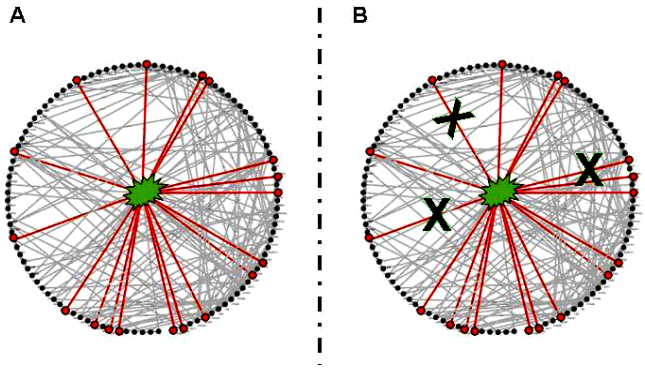

Figure 4. Disruption of host-pathogen interaction as a strategy to control infection. The pathogen inside a host acquires the status of dominant hub, interacting with many host molecules (presumed to be advantageous for the pathogen). Selective disruption of these interactions in a combinatorial manner might provide a fool-proof strategy to control intracellular pathogen survival and growth.

[47–49]. Lipids in host cells also play an important role in regulation and dynamics. There exists a whole range of signaling events and resultant cellular responses that are directly regulated by lipid moieties inside the host cells. However, the status of the host cell lipidome in mycobacteria infected cells remains an unexplored area of investigation. Analysis of the host lipidome under infected condition becomes particularly important since mycobacteria, in addition to modulating the host proteome, are also capable of influencing the host lipidome. Since many of these lipids are directly involved in regulating various biological processes, a detailed analysis of the altered lipid composition in infected host cells becomes a key field of investigation.

Modularity of biological systems is now an accepted phenomenon. Various bio-molecular components constitute macro-modules in the system. Therefore, an integrated analysis of more that one macro-module including proteome, lipidome and metabolome, of both host and the pathogen would provide a far more comprehensive picture of the nature of communication that the two systems engaged in, thereby paving ways for devising new strategies to disrupt this process in favor of the hosts.

7. Developing new drugs and drug targets

There is a continuous need to develop and identify new drug targets. While there are several reasons to do so, the two most important are (a) development of resistance by the pathogen against existing drugs and (b) the likelihood of undesirable side-effects by consumption of the existing drugs owing to the approach through which they were discovered in

the first place. Therefore these two issues become core of all the efforts leading towards identifying and developing new drugs. With the compelling evidence of global regulatory structures (motifs) and networked behavior, focus now is to identify multiple nodes/components in a unique combination that together would constitute the target for a particular diseased state [50,51]. At least as a proof of concept, we have previously demonstrated the participation of multiple nodes as a signature contributing towards various physiological outputs in B lymphoma cells [5]. Here disruption of intracellular signaling components by siRNA-mediated silencing was used to perturb the system and then combined with various experimental and modeling tools to identify the molecular signatures.

A potentially important extension of this approach would be to apply it to conditions where cells/organisms are challenged by pathogens/diseases. Such challenges would represent another kind of perturbation. The vast majority of the pathogens of the human hosts have co-evolved along with the host so that they continue to successfully infect and survive in them. They are able to do so by evading host immune response that results from a complex interplay between host cell machinery and the pathogen. This host-pathogen interplay manifests at every level of cellular regulatory machinery including signaling network, metabolic network and transcriptional regulatory network. Pathogen-derived molecules tend to hijack these regulatory modules and mend them in such a manner that it continues to survive inside the host cell. Considering the complexity of the cellular regulatory network, it becomes necessary for us to first identify points of interaction between molecular components of the pathogen, and the regulatory networks of the host cell. Once we identify these points of invasion (interaction), the sphere of their influence in the regulatory network becomes the subsequent target to establish. This would help to identify both the long- and short-range influences that result from these interactions. It clearly becomes important then to identify all the pathways and regulatory intermediates involved. Tools for large network-level analyses, would then enable us to identify molecular signatures for individual pathogen-specific perturbation conditions. This, in turn, is likely to open avenues for interfering with these interactions of pathogen with the host regulatory network at multiple key points, thus restricting the possibility of any short-circuits to get established by the pathogen when we are targeting these points with specific drugs. Also, since the molecular signatures identified eventually constitute the cellular response axis, by again taking a leaf from our previous studies, we would have the opportunity to delineate how this cellular response axis is manipulated under various pathogenic conditions. An immediate implication of such characterizations would be the identification of the multiple nodes of the regulatory network that tend to be co-opted by the pathogens. These nodes could be identified experimentally, and then validated through mathematical analytical tools. Given that the shift in response axes is mediated by pathogen-derived bio-molecules through their interactions with these target nodes, identification of these interacting partners and designing small molecules that can inhibit these interactions will provide a novel avenue for chemotherapy. The prevention of subjugation of the host cell machinery by an intracellular parasite may suffice to ensure clearance of the pathogen. This is especially likely in the case of infections that target the host macrophage. This cell type is equipped with a potent defense mechanism that is often suppressed by the invading pathogen. Alternatively, such an approach would at least enhance the susceptibility of the pathogen to existing drugs. This approach has potential application for both preventive and curative medicines. Importantly, as the emphasis of this approach is to inhibit molecular interactions between the pathogen and the host, rather than to directly kill the pathogen, it is also less likely to promote the emergence of drug resistant strains of the pathogen. Finally, at least in the case of macrophage-targeting pathogen, a strategy of enabling host cell-mediated elimination of pathogen would also lead to more effective presentation of pathogen-specific antigens by the MHC molecules of the macrophage. The possibility that the potent immune response that would naturally result could in turn provide long-term protection against subsequent infection can, therefore, not be ruled out.

The strategy described above constitutes a novel way of dealing with human infectious diseases, where multiple components of the regulatory system would be targeted in a comprehensive manner. Identification and establishment of these targets would require rigorous experimentation towards both basic understanding of cellular regulatory process as well as perturbation induced cooption/restructuring of the regulatory network. It would then provide us with handles to manipulate specific pathogenic conditions in a more customized manner. In other words these approaches demand constitution of another holistic branch of study, which could be termed as 'systems pharmacology' (figure 5).

Interestingly, the approach also seems to be suited for addressing the second issue raised in the beginning of this section. As we expect to converge to a list of molecular components that could potentially be targeted, there is a strong

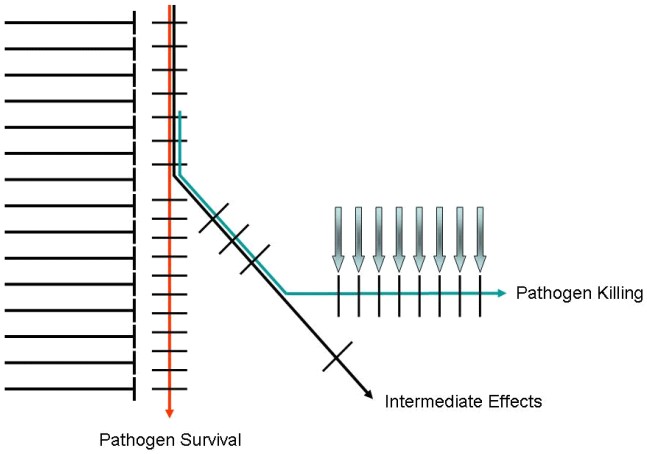

Figure 5. Targeting host factors for controlling pathogen survival. Small black lines represent components of the host classified along the three different axes of pathogen survival, killing and intermediate or no effect. Note the unique and overlapping components along the three axes. Also highlighted is the way various components should be targeted along different axes (i.e., inducing/activating the ones along the killing axis while inhibiting those along the survival axis).

likelihood of synergism between nodal components of the host system. The synergism could well be predicted depending upon the amount of information we can generate for the network as well as acquiring the ability to suitably model them. The models probably then could also inform us as to which of the nodal interventions may give rise to some undesirable outcomes. Additionally, such models would also help to extract a shortlisted set of molecules from the bigger set, which could then be partially targeted in order to get full protection. Combination of nodes can just open the avenues for exploitation in order to get rid two of the most fundamental issues while designing, developing and identifying new drugs.

8. Developing new diagnostics

Since therapeutics and medicine follow the results of diagnostic tests, it becomes extremely important to have a nearly fail-safe system for accurately diagnosing diseased conditions. Unfortunately, current diagnostic techniques, although quite efficient in correlating with diseases, however come to light only after patients have already suffered significantly. And in severely debilitating disease conditions such as cancer, heart disease, and infectious diseases like tuberculosis, the lag between early and late diagnosis makes all the difference between the responses of the patients to the treatment [52]. As we have discussed in the beginning of this article, diagnostic tools suffer almost equally or even more because of the reductionist way of their discovery than the medicine. Under most circumstances, it is our approach that serves as the limitation in identifying the disease conditions at preliminary stages. Majority of the indirect and direct diagnosis depends on one or few highly correlated markers, many of them reach to detectable levels only at the advanced stages of diseases.

With the advancement of new high-throughput techniques, it has become now possible to monitor large numbers of proteins and even scoring for low abundance proteins can be done efficiently. Therefore monitoring global changes in the metabolome, proteome and lipidome is steadily evolving into a popular technique for discovering bio-markers for specific condition. As has been the focus throughout this manuscript, we re-emphasize here the potential of adopting the global systems level approach while developing new diagnostic tools. Thus instead of having a single highly correlated marker for any physiological condition, one can look for patterns across several of the constituent components that could be monitored from the samples obtained. For example, one recent analysis reveals that the human blood leukocyte response to acute systemic inflammation includes the transient dysregulation of leukocyte bioenergetics and modulation of translational machinery [53]. These findings provide insight into the regulation of global leukocyte activities as they relate to innate immune system tolerance and increased susceptibility to infection in humans. So far the most studied and widely explored sample consists of bio-fluid including blood, serum, urine, plasma and cerebrospinal fluid [54]. The advantage of using biofluid lies in the consideration that they represent an ensemble of majority of tissues of the body. A molecular profiling of these biofluids therefore becomes potentially of great prospect in tapping subtle patterns in the constituent, largely proteins and other metabolites, of these biofluids [54]. Identification and scoring for low abundance proteins is extremely important. For example, a person with initial stages of tumor development, or with latently infected tuberculosis, may possess skewed levels of a given constituent in the serum or other body fluids. However, considering that the sample withdrawn is far from the site of actual infection/tumor and the consequent dilution effect in the circulating blood, will bring the levels of the low abundance molecules to even more undetectable levels. Even if they are detected they may not serve the purpose owing to extreme variability in the quantitation, a limitation of biological scoring that gets more pronounced with low abundance molecules. A major leap in the generation and analysis of complex biological data was the transformation of mass-spectrometry from a single peptide/molecule identification/discovery into a tool

that could analyze any kind of complex biological sample containing large number of constituents. In addition to their ability to resolve complex biological samples, this technique can also identify low and extremely low abundance molecules as well in a complex sample. This ability makes mass-spectrometry and related tools (LC-MS-MS, MALDI-ToF-ToF, etc.) as one of the most powerful techniques for developing newer global diagnostic tools. A mass-spectrometric analysis of biofluids like serum or plasma could accurately identify, and in most cases quantify, the majority of its constituent proteins or lipids to high precision [55]. Data sets generated from a sizeable amount of diseased individuals as well as relevant control and healthy individuals could be compared to extract significantly affected molecules specific to the diseased condition in question. One can apply various statistical tools to extract significantly deviating sets of molecules. The majority of the significant list here also would comprise intermediate or high abundance molecules, raising the fear of losing low abundance sets here as well. Therefore, instead of just looking at significant deviations, some lesser significant but consistently affected molecules should also be considered. An interesting approach to classifying various molecular patterns across different sample types would be to follow standard clustering techniques. Thus the majority of co-variating molecules would normally cluster together. When a particular cluster correlates well with a physiological state, this could then together constitute the global marker set for that particular condition. In such cases, it will not be the list of the molecules that constitute the bio-markers, rather the combination of molecules with their integrated quantitative pattern would define a physiological state. With a little more advancement one can catalogue all possible deviations for various stages of a physiological state (for example, various stages of tumor development) and develop an algorithm that can provide a support vector machine platform and accurately predict the physiological state of new data sets generated from blind samples. This could truly transform the bio-marker discovery into diagnostics development approach.

Having said that, the ability of the algorithm generated based on fairly large sets of samples could then actually be applied to two of the most intriguing aspects of medicine, which so far has been enigmatic. The first aspect constitutes predictive and preventive medicine. A global biomolecular profiling of a healthy individual which will predict whether the person will suffer from diabetes, heart disease or a renal carcinoma at the later stage can be generated. This will not only allow the individual to become alert to impending health conditions, for many cases preventive medication could also be started. In other words, we are actually talking about personalized medicines, which undoubtedly would revolutionize the medical practice that we follow currently. Another aspect that will be addressed much more efficiently using such high-throughput diagnostic approach is community medicine. There are several pathogenic and physiological diseases/disorders that are endemic to a particular region. Endemic pathogenic disease depends on two factors, first a relatively high load of causative agent as well as its vector in the locality and, second, a relatively higher susceptibility of the community against the disease, making them more vulnerable. A population wide global profiling in comparison with relevant control population would help us to identify the reasons for the susceptibility of the individuals in a population and probably a cure for the community. In fact, regular sampling and monitoring of these global profiles with efficient mathematical modeling could also be used to accurately predict a disease outbreak in a niche. This constitutes another facet to the advantages of adopting systems biology approach in medicine. However, discussing this any further would fall beyond the scope of this article. Readers are suggested to read [56,57] for further information on this aspect.

9. Challenges and future direction

Undoubtedly systems biology is being considered the answer to majority of the road-blocks that we face today in the field of life sciences as a whole, and medicine in particular. The path from new discovery to diagnostics and development of new medicines, and treatment of many dreaded diseases seems achievable with this approach [58]. However there are several challenges that need to be faced and addressed in order to achieve the goal set for systems biology.

The first and foremost challenge comprises the high-throughput experimental approaches. Several questions have been raised about the variability that is inherent to the high-throughput experimental systems. Thus reproducibility of the results raises severe concern. For example, a simple RNAi screening performed for the same disease by different laboratories has yielded an almost non-overlapping set of molecules important for the progression of that particular disease [42,59,60]. Similarly, mass spectrometric analysis results vary with slight differences in the machine used and condition under which they were performed. A similar fate is also shared by microarray experiments where independent experimental results depend so much on the laboratory where it was performed. To keep a tight vigil on the quality of data set

generated, various standard protocols and information sharing platforms have been established, for example MIAME (Minimum Information About Microarray Experiments) [61] and now on the same line MIARE (Minimum Information About RNAi Experiments) [62,63]. These steps have helped to standardize the experimental conditions across the laboratories around the globe and thus have helped bringing in more consistency in the generated data sets.

Additionally, most of these high-throughput experiments involve extremely sensitive detection techniques, raising the possibility of high noise levels in the generated data sets. Specially, mass-spectrometric data sets (peptides from unknown sources), microarray data (hybridization conditions) and yeast two-hybrid screens for protein-protein interactions (false positives) are bogged with the problem of high noise levels. Stringent statistical analysis becomes an important step to filter out noise from the large scale data generated. Many experiments are repeated multiple times preferably with reagents from different sources in order to ascertain the validity of the results obtained.

The management of huge data sets generated by these experiments constitutes another level of challenge. A researcher trained in biology is most likely to find it difficult managing, leave apart the analysis, of the genome or proteome level data sets generated. With stringent statistical analysis as an integral step, one can begin to realize how biology is evolving into the multi-disciplinary field of systems biology. Similarly, integration of data sets into effective mathematical modeling requires inputs from physicists as well as mathematicians and computational experts, making it a truly global subject. However, this brings in another challenge since no common language yet exists between a biologist and a mathematician or a physicist. Thus exchange of information on a regular basis among collaborators from multiple disciplines has become the necessity for the emerging field of systems biology. The corresponding evolution of the academic network, as a result of the development of systems biology is depicted in figure 1.

While the challenges discussed here are staggering they have, nonetheless, paved ways for more innovative thinking among biologists. The focus has shifted towards out-of-the-box thinking, taking clues from totally unrelated fields like engineering, social sciences, statistical physics and computer sciences and applying them to biological systems to obtain a more comprehensive perspective of the system. At the same time we also expect to see large, ambitious projects aimed at generating high-throughput data sets enabling us to model either the whole cell, the whole organ, the organ system, or even the whole human body. In fact we already have whole cell simulators available online now such as virtual cell and e-cell. Many experimental data sets can be fed into these virtual cells and resulting regulatory events could be resolved through in-silico simulations. With ever expanding data sets arising and being incorporated into such whole cell simulators, we would probably achieve our target of throwing in any small molecule compounds into the cells in-silico and explore what the cell will do under a particular set of conditions. People have also modeled, though at a very primitive scale, the response of the immune cell network to external agents like pathogenic bacteria [64]. While there remain several unknown parameters in such modeling exercises they, nonetheless, provide new dimensions into the analysis and understanding of biological system. These models could further be improved iteratively with incorporation of more dense data sets, enabling us to predict very precisely the outcome of various biological phenomena such as infections, malfunctions and treatments. Even more intensive and extremely high resolution data sets could eventually lead us to a model for the whole multicellular organism with all the features of the organism incorporated into the model. The advantages are obvious, right from understanding the system, to designing better ways of clinical trials, as well as devising newer diagnostic tools. In short, systems biology promises to revolutionize the field of medicine eventually fulfilling the long cherished desire and now seriously needed milestone of predictive, preventive and personalized medicine.

References

[1] Ahn A C, Tewari M, Poon C S, *et al* 2006 *PLoS Med* **3** e208.
[2] Ahn A C, Tewari M, Poon C S, *et al* 2006 *PLoS Med* **3** e209.
[3] Barabasi A L and Oltvai Z N 2004 *Nat. Rev. Genet.* **5** 101.
[4] Kumar D, Dua R, Srikanth R, *et al* 2008 *BMC Res Notes* **1** 81.
[5] Kumar D, Srikanth R, Ahlfors H, *et al* 2007 *Mol. Syst. Biol.* **3** 150.
[6] Cohen A A, Geva-Zatorsky N, Eden E, *et al* 2008 *Science* **322** 1511.
[7] Zhang D H, Salto-Tellez M, Chiu L L, *et al* 2003 *Life Sci.* **73** 3189.
[8] Tailleux L, Waddell S J, Pelizzola M, *et al* 2008 *PLoS ONE* **3** e1403.
[9] Ding X R, Yang J, Sun D C, *et al* 2008 *Pharmacogenomics J* **8** 61.
[10] Fink J, Gu F, Ling L, *et al* 2007 *PLoS Negl Trop. Dis.* **1** e86.
[11] Tan S L, Ganji G, Paeper B, *et al* 2007 *Nat. Biotechnol.* **25** 1383.
[12] Yildirim M A, Goh K I, Cusick M E, *et al* 2007 *Nat. Biotechnol.* **25** 1119.

[13] 2004 *Nature* **431** 931.
[14] Collins F S and McKusick V A 2001 *Jama* **285** 540.
[15] Strogatz S H 2001 *Nature* **410** 268.
[16] Alon U 2007 *Nat. Rev. Genet.* **8** 450.
[17] Voit E, Neves A R and Santos H 2006 *Proc. Natl. Acad. Sci.* USA **103** 9452.
[18] Kirschner M W 2005 *Cell.* **121** 503.
[19] Watts D J and Strogatz S H 1998 *Nature* **393** 440.
[20] Jeong H, Mason S P, Barabasi A L, *et al* 2001 *Nature* **411** 41.
[21] Bar-Yam Y and Epstein I R 2004 *Proc. Natl. Acad. Sci.* USA **101** 4341.
[22] Guan Y, Myers C L, Lu R, *et al* 2008 *PLoS Comput. Biol.* **4** e1000165.
[23] Leclerc R D 2008 *Mol. Syst. Biol.* **4** 213.
[24] Albert R, Jeong H and Barabasi A L 2000 *Nature* **406** 378.
[25] Jeong H, Tombor B, Albert R, *et al* 2000 *Nature* **407** 651.
[26] Cohen R, Erez K, ben-Avraham D, *et al* 2000 *Phys. Rev. Lett.* **85** 4626.
[27] Callaway D S, Newman M E, Strogatz S H, *et al* 2000 *Phys. Rev. Lett.* **85** 5468.
[28] Barabasi A L and Albert R 1999 *Science* **286** 509.
[29] Haldar S, Jena N and Croce C M 1995 *Proc. Natl. Acad. Sci.* USA **92** 4507.
[30] Ishikawa Y, Kusaka E, Enokido Y, *et al* 2003 *Mol. Cell. Neurosci.* **24** 451.
[31] Papin J A, Hunter T, Palsson B O, *et al* 2005 *Nat. Rev. Mol. Cell. Biol.* **6** 99.
[32] Zha J, Harada H, Yang E, *et al* 1996 *Cell.* **87** 619.
[33] Flynn J L and Chan J 2001 *Annu. Rev. Immunol.* **19** 93.
[34] Balcewicz-Sablinska M K, Keane J, Kornfeld H, *et al* 1998 *J. Immunol.* **161** 2636.
[35] Blander J M and Medzhitov R 2004 *Science* **304** 1014.
[36] Blaser M J and Kirschner D 2007 *Nature* **449** 843.
[37] Clemens D L, Lee B Y and Horwitz M A 2002 *Infect. Immun.* **70** 5800.
[38] Warner D F and Mizrahi V 2007 *Nat. Med.* **13** 282.
[39] Lamichhane G and Bishai W 2007 *Nat. Med.* **13** 280.
[40] Kuijl C, Savage N D, Marsman M, *et al* 2007 *Nature* **450** 725.
[41] Jayachandran R, Sundaramurthy V, Combaluzier B, *et al* 2007 *Cell.* **130** 37.
[42] Brass A L, Dykxhoorn D M, Benita Y, *et al* 2008 *Science* **319** 921.
[43] Prudencio M, Rodrigues C D, Hannus M, *et al* 2008 *PLoS Pathog.* **4** e1000201.
[44] Schlabach M R, Luo J, Solimini N L, *et al* 2008 *Science* **319** 620.
[45] Silva J M, Marran K, Parker J S, *et al* 2008 *Science* **319** 617.
[46] Krishnan M N, Ng A, Sukumaran B, *et al* 2008 *Nature* **455** 242.
[47] Fratti R A, Chua J, Vergne I, *et al* 2003 *Proc. Natl. Acad. Sci.* USA **100** 5437.
[48] Gehring A J, Rojas R E, Canaday D H, *et al* 2003 *Infect. Immun.* **71** 4487.
[49] Ehrt S and Schnappinger D 2007 *Nat. Med.* **13** 284.
[50] Ma'ayan A, Jenkins S L, Neves S, *et al* 2005 *Science* **309** 1078.
[51] Prill R J, Iglesias P A and Levchenko A 2005 *PLoS Biol.* **3** e343.
[52] Anderson N L 2005 *Mol. Cell. Proteomics.* **4** 1441.
[53] Calvano S E, Xiao W, Richards D R, *et al* 2005 *Nature* **437** 1032.
[54] Veenstra T D, Conrads T P, Hood B L, *et al* 2005 *Mol. Cell. Proteomics.* **4** 409.
[55] Hanash S M, Pitteri S J and Faca V M 2008 *Nature* **452** 571.
[56] Eubank S, Guclu H, Kumar V S, *et al* 2004 *Nature* **429** 180.
[57] Reis B Y, Kohane I S and Mandl K D 2007 *PLoS Med.* **4** e210.
[58] Henney A and Superti-Furga G 2008 *Nature* **455** 730.
[59] Goff S P 2008 *Cell.* **135** 417.
[60] Konig R, Zhou Y, Elleder D, *et al* 2008 *Cell.* **135** 49.
[61] Brazma A, Hingamp P, Quackenbush J, *et al* 2001 *Nat. Genet.* **29** 365.
[62] Brazma A, Krestyaninova M and Sarkans U 2006 *Nat. Rev. Genet.* **7** 593.
[63] Boutros M and Ahringer J 2008 *Nat. Rev. Genet.* **9** 554.
[64] Thakar J, Pilione M, Kirimanjeswara G, *et al* 2007 *PLoS Comput. Biol.* **3** e109.